Communities within the statistics blog network

John Johnson

Social Network Analysis Class on Coursera

November 9, 2012

Introduction

The purpose of this project was to explore the structure of the statistics "blogosphere" (loosely, the collection of statistics weblogs) and its relationship to content. The following can be considered as a preliminary, hypothesis generating analysis.

The analysis compares two ways of clustering blogs that discuss statistics: through community detection considering the statistics blogosphere as a network of links defined by their blogrolls, and through unsupervised learning techniques based on the content of the blogs.

Data collection

Identification of blogs

Statistics blogs were identified as follows:

- All blogs contributing to the aggregator site "Statblogs.com" were included as "seed" blogs.
- For all the seed blogs, the blogrolls were manually reviewed for links to other blogs, and also new blogs.
- New blogs were evaluated for whether their content was mostly statistical. If so, they were included in the analysis and had their blogrolls analyzed.
- The process above was continued until no more statistics blogs were found.

Each blog was represented as a node in the network.

Links

Links between blogs were constructed on a manual basis by reviewing blogrolls. This was because blogrolls took different forms on different blogs. Many blogs had blogrolls in sidebars, but had different titles identifying them. Some blogs, such as Andrew Gelman's blog, had a blogroll on an information page rather than on the front page. Many blogs had a blogroll that included non-blog sites, such as links to societies, papers, or other information. In addition, Andrew Gelman's blog has moved a couple of times and is currently located at http://www.andrewgelman.com. References to the old blog were moved to this location. Other blogs, such as http://simplystatistics.org, were treated similarly. Finally,

distinctions between uninteresting differences such as http://www.url.com and http://url.com were resolved into the same url. Each link was represented as a directed edge from the node corresponding to the blog posting the blogroll to the node corresponding to the target of the link.

Other efforts to harvest links automatically ended in non-interesting and clearly incorrect results. For instance, links were harvested from the blog entries, and also by scraping the front page of each blog. These efforts were subsequently abandoned.

Content

For each blog, the feed was found by using href in a link tag of type RSS or Atom. The feed was accessed by using FeedParser and saved into Javascript Object Notation (json) files. These files stored the title, links, and content. HTML tags were removed from the content before further analysis.

Data storage

For ease of collection, blogs were stored as edge lists in a text file with URL as keys. This list is included as an appendix to this document. This file was read by a custom Python script into NetworkX graph. From here, the blog titles were added as labels. NetworkX was used to write a Graph Markup Language (GML) file, which could be read into Gephi.

Analysis

Network analysis

Network analysis will consist of the following analyses:

1. Centrality, betweenness, and PageRank analyses
2. Community detection, based on modularity

In addition, centrality measures will be compared to Erdős-Renyi graph with the same number of nodes and edges.

Content analysis

Content from the json files was converted into a term document frequency matrix. A term document frequency matrix has all words in all the blogs (excepting “a,” “an”, and “the”) marking the columns and all blogs marking the rows. An entry in this matrix has the number of times the word in the column appears in the blog in the row. The Pearson correlation statistic between two rows was used to determine the similarity in content between two blogs. This statistic ranges from -1 to 1 and corrects for the length of the posts while determining if similar words are used.

The k-means algorithm is an unsupervised learning algorithm that divides the blogs into clusters. In other words, the algorithm finds groups of blogs that are all similar to each other based on content. One important input into this algorithm is the number of clusters. The input here will be the number of communities above, to see if the communities detected by modularity are similar to those detected by k-means on content.

The cluster id will be included in the gml file so that Gephi can show visualizations based on cluster id.

Combined analyses

The assignments to communities from the community detection algorithm will be compared to the assignment to clusters using chi-squared test with the following hypotheses:

Null: there is no relationship between assignment to community and assignment to cluster based on content

Alternative: there is a relationship between assignment to community and assignment to cluster based on content

Visualizations using two separate versions of coloring the graph (one based on content cluster and the other based on community) will be used to support this analysis.

Results

Basic graph properties

The resulting graph is a directed graph with 528 nodes (blogs) and 786 edges (directed links). There are 16 connected components, with the giant component accounting for 507 (96%) of the nodes and 780 (99%) of the edges. The graph is displayed in Figure 1: Graph of Statistics Blogosophere.

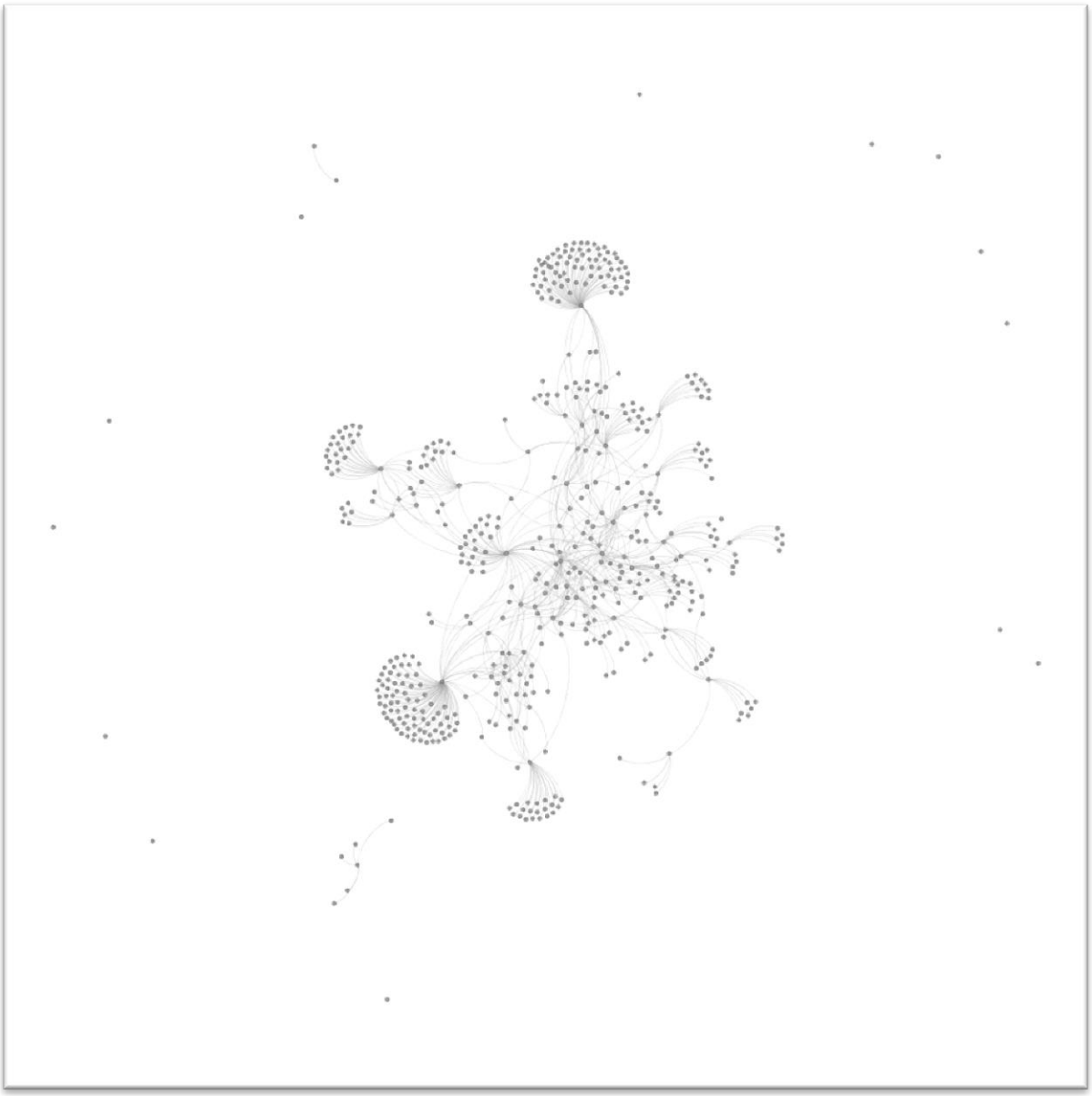

Figure 1: Graph of Statistics Blogosphere

Degree and Centrality

The 5 most linked-to blogs (i.e. with highest in-degree) are as follows:

- R-Bloggers (aggregator site)
- Andrew Gelman
- Stats Blogs (aggregator site)
- John D. Cook
- Flowing Data

The 5 blogs that link to the most blogs (i.e. with highest out-degree), not counting R-Bloggers and Stats Blogs (the two aggregator sites) are as follows:

- Nuit Blanche
- Cool Infographics
- Hong Lang Wang
- Omics! Omics!
- Junk Charts

The 5 most central blogs according to the PageRank algorithm are as follows:

- R-Bloggers
- Andrew Gelman
- Stats Blogs
- Terry Tao
- Psychological Statistics

The 5 blogs closest to others according to path length (with undirected edges) are as follows:

- My Biased Coin
- Brain Stat
- Society to Supress the Correlation Coefficient
- Code and Culture
- Serious Stats

The 5 blogs that are most likely to lie on the shortest paths between others (edges counted as undirected) are as follows:

- Andrew Gelman
- Normal Deviate (Larry Wasserman)
- Nuit Blanche
- Junk Charts
- Stats Chat

Comparison to Erdős-Renyi graphs

From the visualization, the blogs graph is qualitatively different from a random graph.

The distribution of betweenness centrality is clearly different from that of a random graph, as shown in Figure 2: Betweenness centrality distribution of statistics blogs graph and Figure 3: Betweenness centrality distribution of random graph.

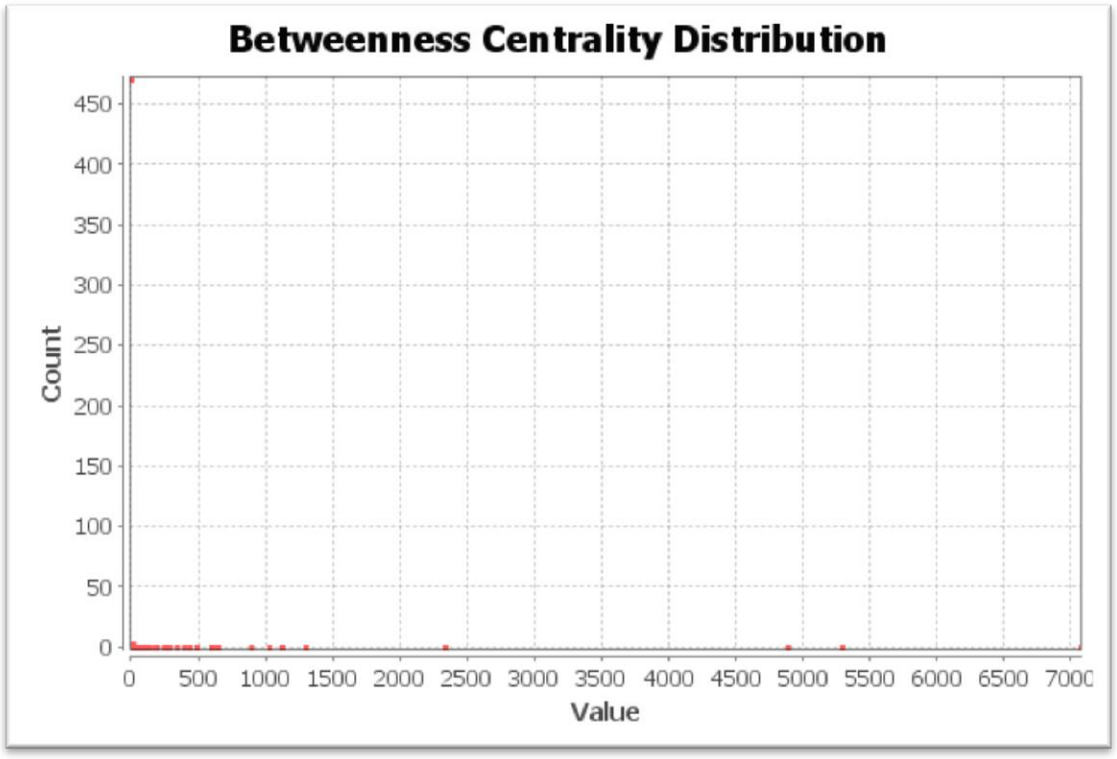

Figure 2: Betweenness centrality distribution of statistics blogs graph

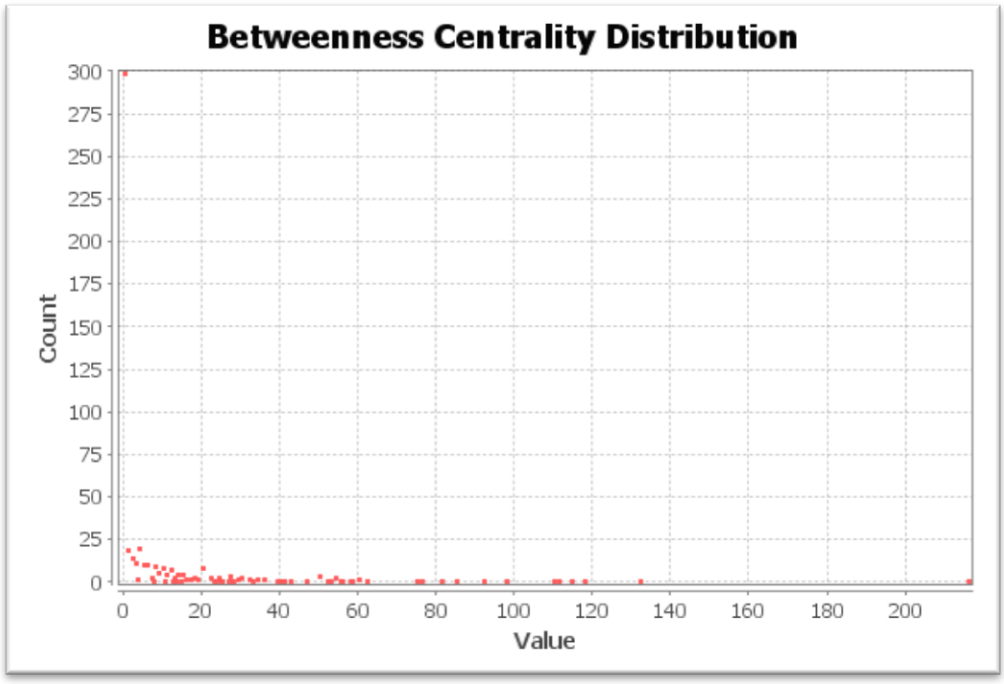

Figure 3: Betweenness centrality distribution of random graph

The PageRank distribution of the statistics blog graph is given in Figure 4: PageRank distribution of statistics blogs graph and the distribution for the random graph given in Figure 5: PageRank distribution of random graph.

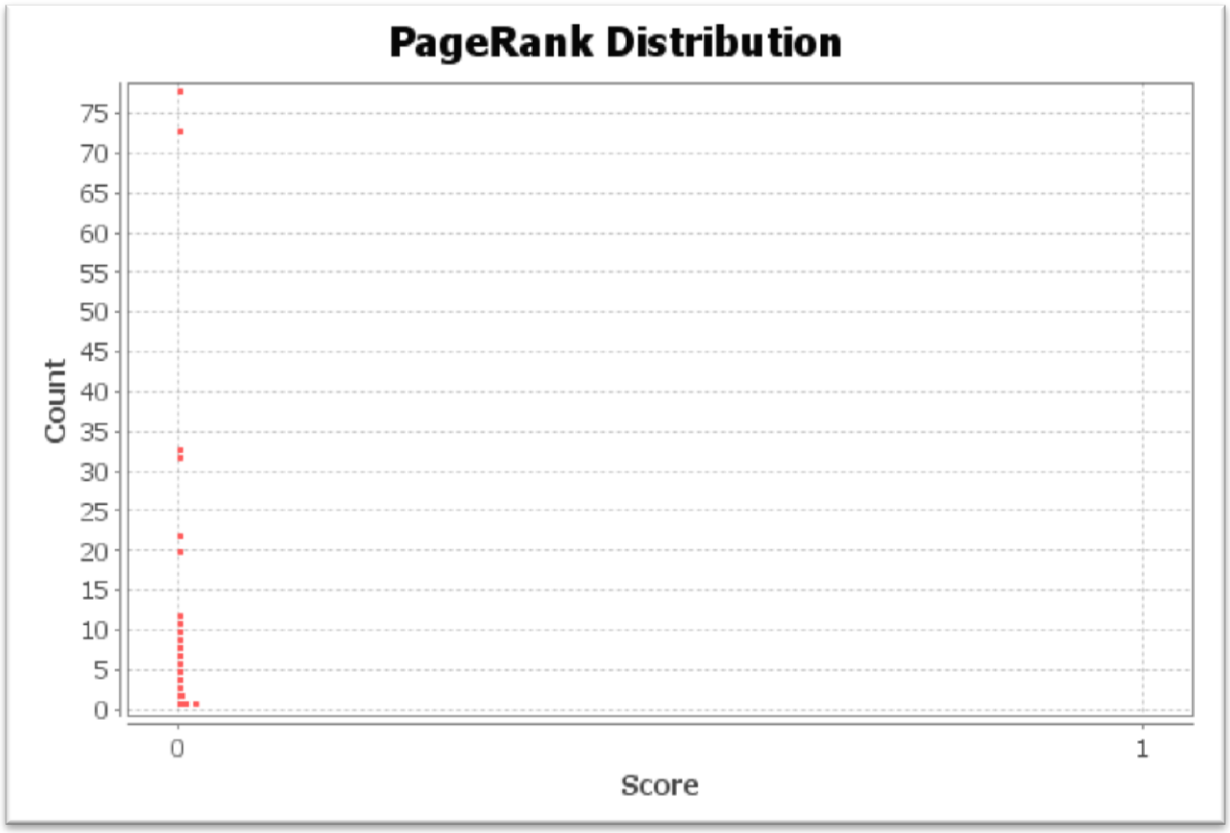

Figure 4: PageRank distribution of statistics blogs graph

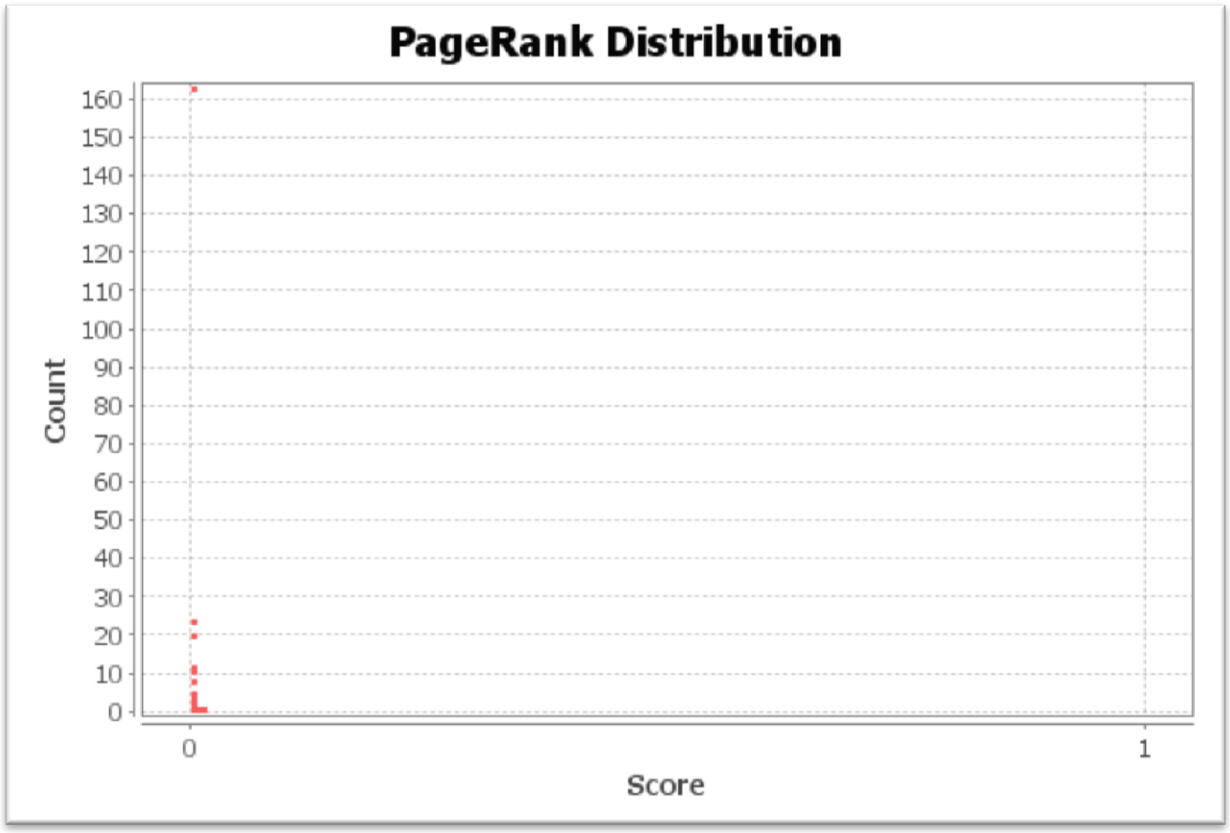

Figure 5: PageRank distribution of random graph

The following are comparisons of the statistics graph with random graphs

Item	Statistics blogs	Random graph
Diameter	8	9
Average path length	3.84	2.83
Number of shortest paths	2919	11061
Average degree	1.49	1.5

Modularity and clustering

The community detection algorithm based on modularity is shown in Figure 6. The different colors for the nodes indicate the different communities.

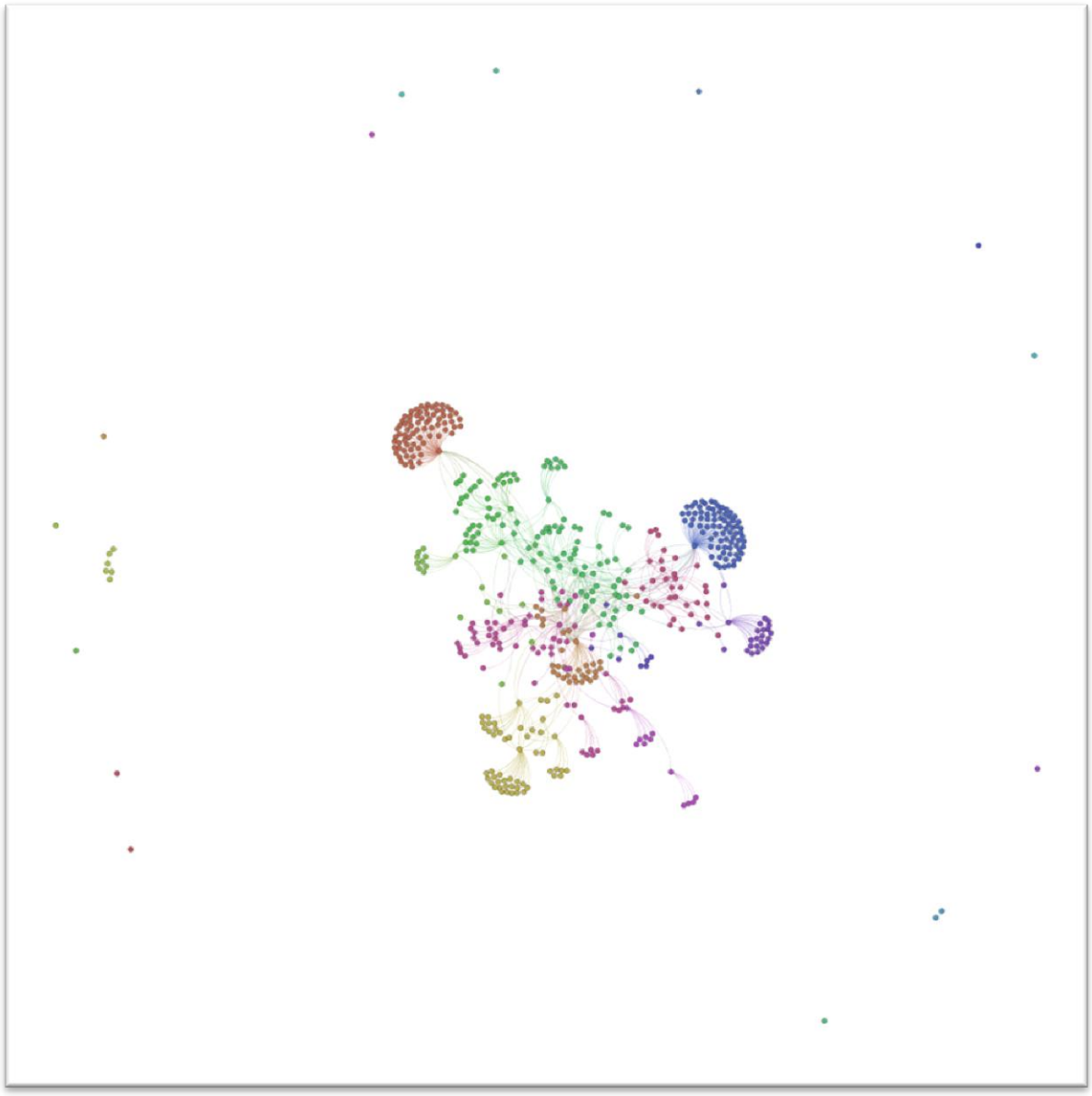

Figure 6: Statistics blogs, colored by community based on modularity

In Figure 7, the graph with nodes colored based on content (clustered using the k-means algorithm) is shown. About 65% of the blogs yielded analyzable text. (Errors may have occurred due to timeout errors, inability to find a feed, or other internet glitches.) Many of the communities are in the same cluster based on content as well, and the chi-square test for the relationship between community based on social network and clustering based on content is <0.001, indicating evidence of a relationship.

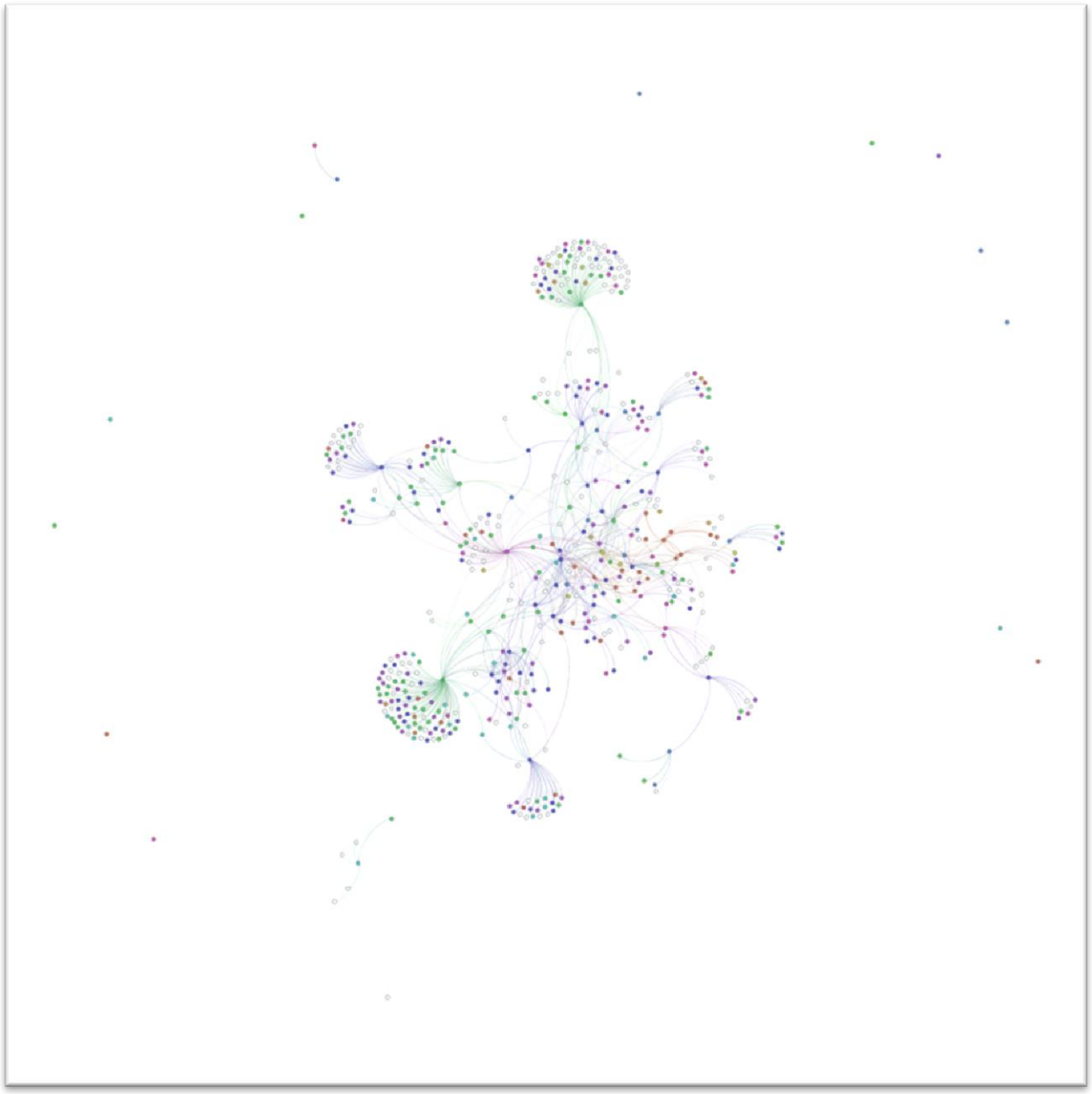

Figure 7: Blogs graph colored based on content

Discussion

Interpretation

Diameter and average degree were very similar between the random graph and the statistics blogs graph. However, the distributions of centrality were very different. The statistics blogs had several nodes with very high centrality and a very large number of nodes with very low centrality. The PageRank distribution was similar – just a very few nodes with high PageRank (0.008 and higher) and most with a

low PageRank. By contrast, the random graph showed a much more graduated effect. These centrality distributions reflect the notion that the statistics blogosphere has a central set of blogs that talk about general statistical content and a large number of subcommunities. Indeed, the blogs with the highest centrality measures were the two aggregator sites (R-bloggers and Stats blogs) and Andrew Gelman's blog. Finally, the number of shortest paths reflects the notion that a lot of the subcommunities link back to the statistics blogs through one or two sites. For instance, a large number of sites are can be discovered from the statistics blogs through the Nuit Blanche blog, as shown in Figure 8.

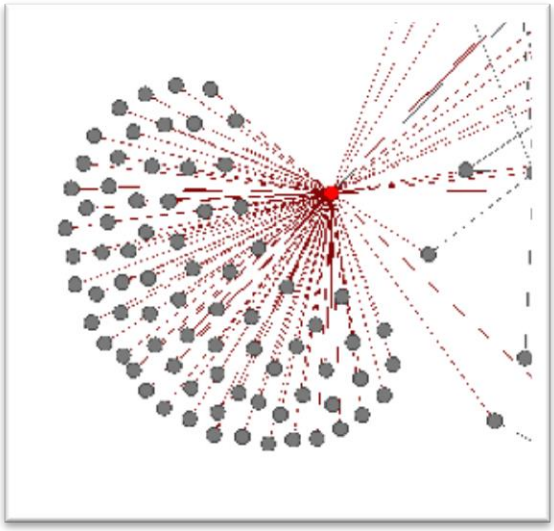

Figure 8: Portion of statistics blogs, with Nuit Blanche colored in red. Nuit Blanche serves as the link between many blogs and the rest of the statistics blogs.

The community detection algorithm seemed to detect many, but not all, the subcommunities that are of the nature shown in Figure 8.

Even given the apparent relationship between community and content, it is clear there is a diversity of content within each community, as shown qualitatively in Figure 7.

Limitations and Extensions

The identification of blogs and links between blogs was done by one person over the course of three days. Some subjective assessments were made about the topics of the blogs. This process can be improved by having multiple people do the manual work and coming to a consensus.

Several extensions can be used to make this analysis more robust. Many words are used uniquely and therefore may be uninteresting in the analysis (and may result in an unnecessarily larger computing time). Some common words may not be useful at all (although the most common words "a," "an," and "the" were removed). These words may be removed or down weighted in a more sophisticated analysis:

- A larger stopword (uninteresting word) list can be used
- The words that are present in, say, 20%-80% of the blogs can be used, with the other words discarded

- Common words can be downweighted, such as through the use of the TF-IDF (Term Frequency * inverse document frequency) approach. (This approach was attempted, but appeared to give uninteresting results.)

In addition, there are other clustering methods that can be tried:

- Topic discovery using non-negative matrix factorization on the term document frequency matrix (Segaran, 2007)
- Named entity extraction

Appendix

Tools

- Python: http://www.python.org for general programming and analysis
- NetworkX: http://networkx.lanl.gov for converting and analyzing graphs
- NLTK: http://www.nltk.org for analyzing content
- Gephi http://www.gephi.org for visualizing and analyzing graphs
- FeedParser: http://code.google.com/p/feedparser/ for extracting blog content

Data

The following is an edge list of the network, with blogs keyed by URL. The format of each line is blog;link1,link2,etc.

```
http://stataccess.blogspot.com;http://www.r-
bloggers.com/,http://www.guardian.co.uk/news/datablog,http://blogs.sas.com
/content/sascom,http://www.flowingdata.com,http://radar.oreilly.com,http:/
/dailytekk.com,http://googlepolitics.blogspot.com,http://utvbloggen.se,htt
p://swedeneurostat.blogspot.com,http://www.perceptualedge.com/blog,http://
www.cros-
portal.eu/frontpage,https://gephi.org,http://www.informationisbeautiful.ne
t,http://blogstats.wordpress.com,http://hdr.undp.org
http://adventuresinr.wordpress.com;
http://allthingsr.blogspot.com;
http://alstatr.blogspot.com;http://sinama.org,http://iitstatwizards.blogsp
ot.com,http://iitstatwizards.weebly.com,http://addictedtor.free.fr/graphiq
ues,http://rgm2.lab.nig.ac.jp/RGM2/images.php?show=all&pageID=1405
http://bayesianbiologist.com;http://ubseblz.wordpress.com,http://allendown
ey.blogspot.com,http://www.r-
bloggers.com,http://blog.revolutionanalytics.com,http://www.andrewgelman.c
om,http://www.statsblogs.com,http://www.johndcook.com/blog,http://timharfo
rd.com/articles/undercovereconomist,http://understandinguncertainty.org
http://biostatmatt.com;
http://blogstats.wordpress.com;
http://rdatamining.wordpress.com;
http://botthoughts.wordpress.com;http://trinkerrstuff.wordpress.com,http:/
/www.statsblogs.com
```

http://brainchronicle.blogspot.com;http://www.r-
bloggers.com/,http://www.technologyreview.com/stream,http://infosthetics.c
om,http://gettinggeneticsdone.blogspot.com,http://www.genomesunzipped.org
http://www.carlislerainey.com;
http://christophergandrud.blogspot.com;http://baselinescenario.com,http://
blogs.reuters.com/felix-
salmon,http://ftalphaville.ft.com,http://www.globalpolicyjournal.com,http:
//www.r-
bloggers.com,http://statsblogs.com,http://www.johndcook.com/blog,http://ww
w.themonkeycage.org,http://www.andrewgelman.com/blog
http://darrenjw.wordpress.com;
http://blog.data-miners.com;
http://www.data-mining-blog.com;http://www.datamining-blog.de
http://datamining.typepad.com/data_mining;
http://www.statsmakemecry.com/smmctheblog;http://faculty.chass.ncsu.edu/ga
rson/PA765/statnote.htm,http://davidakenny.net,http://davidmlane.com/hyper
stat/index.html,http://www.r-
bloggers.com,http://www.statsblogs.com,http://cooldata.wordpress.com,http:
//www.spsstools.net,http://www.thejuliagroup.com/blog,http://www.theanalys
isfactor.com
http://doingbayesiandataanalysis.blogspot.com;
http://eagereyes.org;http://fellinlovewithdata.com,http://thewhyaxis.info,
http://junkcharts.typepad.com,http://www.zcliu.org/blog,http://blog.visual
.ly,http://dataremixed.com,http://vizwiz.blogspot.com,http://peltiertech.c
om/WordPress,http://www.excelcharts.com/blog,http://substratumseries.com,h
ttp://thedailyviz.com,http://www.flowingdata.com,http://infosthetics.com
http://blogs.forbes.com/naomirobbins;
http://joelcadwell.blogspot.com;http://www.r-
bloggers.com,http://www.statsblogs.com
http://errorstatistics.com;
http://blog.vinux.in;
http://fishyoperations.com;http://www.r-
bloggers.com,http://www.statsblogs.com
http://gettinggeneticsdone.blogspot.com;
http://gianlubaio.blogspot.com;http://xianblog.wordpress.com,http://unders
tandinguncertainty.org,http://www.andrewgelman.com,http://www.statsblogs.c
om,http://www.r-
bloggers.com,http://normaldeviate.wordpress.com,http://scharrheds.blogspot
.com,http://martynplummer.wordpress.com
http://www.graphoftheweek.org;http://www.statsblogs.com,http://www.r-
bloggers.com,http://datajournalists.org,http://rhodestales.com
http://honglangwang.wordpress.com;http://www.statsblogs.com,http://www.r-
bloggers.com,http://www.inside-
r.org,http://normaldeviate.wordpress.com,http://www.rdatamining.com,http:/
/guidetodatamining.com,http://brenocon.com/blog,http://www.analyticbridge.
com,http://bayesianbiologist.com,http://brainstat.blogspot.com,http://rbal
tman.wordpress.com,http://chrisladroue.com,http://dahuasky.wordpress.com,h
ttp://ivory.idyll.org/blog,http://darrenjw.wordpress.com,http://scienceblo
gs.com/digitalbio,http://dirk.eddelbuettel.com,http://blog.echen.me,http:/
/errorstatistics.com,http://www.flowingdata.com,http://gettinggeneticsdone
.blogspot.com,http://hao1990.blogspot.com,http://heuristically.wordpress.c
om,http://www.homolog.us/blogs,http://jeromyanglim.blogspot.com,http://www
.cnblogs.com/jerrylead,https://wiki.hpcc.msu.edu/display/~johnj@msu.edu,ht
tp://www.johnmyleswhite.com,http://linbaba.wordpress.com,http://taoshistat

.wordpress.com,http://kevin-
gattaca.blogspot.com,http://nextgenseq.blogspot.com,http://ngs-
brescia.blogspot.it,http://oelemento.wordpress.com,http://omicsomics.blogs
pot.com,http://blog.goldenhelix.com,http://pairach.com,http://www.politige
nomics.com,http://polylogblog.wordpress.com,http://polylogblog.wordpress.c
om,http://www.quantumforest.com,http://statmethods.wordpress.com,http://rt
utorialseries.blogspot.com,http://realizationsinbiostatistics.blogspot.com
,http://web.ece.rice.edu/richb,http://www.andrewgelman.com,http://robjhynd
man.com,http://simplystatistics.org,http://statisfaction.wordpress.com,htt
p://statmethods.wordpress.com,http://www.johndcook.com/blog,http://statist
icsforum.wordpress.com,http://lamages.blogspot.com,http://www.cscs.umich.e
du/~crshalizi/weblog,http://understandinguncertainty.org,http://wmbriggs.c
om/blog,http://xianblog.wordpress.com,http://yihui.name/en
http://infosthetics.com;
http://junkcharts.typepad.com/junk_charts;http://www.andrewgelman.com,http
://www.guardian.co.uk/news/datablog,http://peltiertech.com/WordPress,http:
//www.excelcharts.com/blog,http://www.juiceanalytics.com/writing,http://ww
w.theusrus.de/blog,http://www.theusrus.de/blog,http://eagereyes.org,http:/
/www.flowingdata.com,http://www.perceptualedge.com/blog,http://visualmetho
ds.blogspot.com,http://www.joeparry.com/blog,http://designnotes.info,http:
//infosthetics.com,http://www.informationisbeautiful.net,http://www.visual
complexity.com/vc/blog,http://www.maxgadney.com,http://www.coolinfographic
s.com,http://www.sankey-
diagrams.com,http://strangemaps.wordpress.com,http://chandoo.org/wp,http:/
/www.databison.com,http://www.r-
bloggers.com,http://learnr.wordpress.com,http://www.smartdraw.com,http://w
ww.andypope.info/index.htm
http://learnandteachstatistics.wordpress.com;http://lovestats.wordpress.co
m,http://simplystatistics.org,http://www.statschat.org.nz
http://taoshistat.wordpress.com;http://www.andrewgelman.com,http://www.sta
tsblogs.com,http://thelogcabin.wordpress.com,http://understandinguncertain
ty.org,http://xianblog.wordpress.com
http://lamages.blogspot.com;
http://normaldeviate.wordpress.com;http://www.andrewgelman.com,http://xian
blog.wordpress.com,http://www.cscs.umich.edu/~crshalizi/weblog,http://erro
rstatistics.com,http://hunch.net,http://www.mii.ucla.edu/causality,http://
nuit-blanche.blogspot.it
http://junkcharts.typepad.com/numbersruleyourworld;http://www.andrewgelman
.com,http://xianblog.wordpress.com,http://understandinguncertainty.org,htt
p://www.johndcook.com/blog,http://mat.gsia.cmu.edu/blog,http://fivethirtye
ight.blogs.nytimes.com,http://realizationsinbiostatistics.blogspot.com,htt
p://www.iq.harvard.edu/blog/sss,http://thestatsblog.wordpress.com,http://b
logs.wsj.com/numbersguy,http://blogs.sas.com/blognormal,http://datamining.
typepad.com,http://www.statschat.org.nz
http://onertipaday.blogspot.com;http://www.r-
bloggers.com,http://blog.revolutionanalytics.com,http://zoonek2.free.fr/UN
IX/48_R/all.html,http://davegiles.blogspot.com,http://www.flowingdata.com,
http://romainfrancois.blog.free.fr,http://ryouready.wordpress.com,http://l
earnr.wordpress.com,http://weitaiyun.blogspot.com,http://sas-and-
r.blogspot.com,http://rdataviz.wordpress.com,http://blog.diegovalle.net,ht
tp://www.dataminingblog.com
http://allendowney.blogspot.com;
http://onbiostatistics.blogspot.com;http://realizationsinbiostatistics.blo
gspot.com

http://psychologicalstatistics.blogspot.com;http://seriousstats.wordpress.com,http://www.r-bloggers.com,http://societytosupressthecorrelationcoefficient.wordpress.com,http://bps-msc.blogspot.com,http://www.andrewgelman.com
http://www.quantumforest.com;http://codeandculture.wordpress.com,http://errorstatistics.blogspot.com,http://freakonometrics.blog.free.fr,http://ggorjan.blogspot.com,http://allendowney.blogspot.com,http://www.r-bloggers.com,http://www.statschat.org.nz,http://www.johndcook.com/blog,http://www.cscs.umich.edu/~crshalizi/weblog,http://xianblog.wordpress.com
http://rsnippets.blogspot.com;http://bogumilkaminski.home.pl/index.php,http://www.r-bloggers.com,http://www.statsblogs.com
http://www.r-tutor.com;
http://viksalgorithms.blogspot.com;http://www.r-bloggers.com
http://r4stats.com;http://www.r-bloggers.com,http://statsblogs.com
http://rbresearch.wordpress.com;http://www.r-bloggers.com
http://realizationsinbiostatistics.blogspot.com;http://www.ask-cato.com,http://onbiostatistics.blogspot.com,http://blog.revolutionanalytics.com,http://www.r-bloggers.com,http://sas-and-r.blogspot.com,http://onertipaday.blogspot.com,http://www.r-chart.com,http://www.programmingr.com/taxonomy/term/14/0,http://www.cerebralmastication.com,http://learnr.wordpress.com,http://www.statsblogs.com,http://blogs.sas.com/content/iml,http://blogs.sas.com/content/sasdummy,http://blogs.forbes.com/naomirobbins,http://statisticsforum.wordpress.com,http://blog.360.yahoo.com/blog-fjmXwIAjcKjZxsEcaDFYdQ--;_ylt=Ap.10_j8skQEJnU3JvX7hhiqAOJ3?cq=1,http://www.flowingdata.com,http://learnbayes.blogspot.com,http://www.johndcook.com/blog,http://liesandstats.wordpress.com,http://www.mathepi.com/index.html,http://radfordneal.wordpress.com,http://www.andrewgelman.com
http://robjhyndman.com;
http://sas-and-r.blogspot.com;
http://www.sas-programming.com;http://www.r-bloggers.com,http://www.statsblogs.com
http://seriousstats.wordpress.com;http://psychologicalstatistics.blogspot.com,http://www.r-bloggers.com,http://societytosupressthecorrelationcoefficient.wordpress.com
http://sharpstatistics.co.uk;
http://simplystatistics.org;
http://statbandit.wordpress.com;http://www.andrewgelman.com,http://xianblog.wordpress.com,http://www.r-bloggers.com,http://simplystatistics.org,http://www.johndcook.com/blog,http://www.walkingrandomly.com
http://www.statalgo.com;
http://statisfaction.wordpress.com;http://xianblog.wordpress.com,http://freakonometrics.blog.free.fr,http://www.r-bloggers.com,http://robinryder.wordpress.com,http://www.statsblogs.com
http://www.andrewgelman.com;http://xianblog.wordpress.com,http://www.themonkeycage.org,http://statisticsforum.wordpress.com,http://nuit-blanche.blogspot.com,http://www.statsblogs.com,http://simplystatistics.org,http://normaldeviate.wordpress.com,http://www.cscs.umich.edu/~crshalizi/weblog,http://junkcharts.typepad.com/numbersruleyourworld,http://observationalepidemiology.blogspot.com,http://www.johndcook.com/blog,http://messymatters.com,http://www.r-bloggers.com,http://blogs.wsj.com/numbersguy,http://junkcharts.typepad.com

,http://infosthetics.com,http://strangemaps.wordpress.com,http://eagereyes
.org,http://chartsnthings.tumblr.com,http://fivethirtyeight.blogs.nytimes.
com,http://www.decisionsciencenews.com
http://systematicinvestor.wordpress.com;http://cssanalytics.wordpress.com,
http://www.r-bloggers.com,http://engineering-
returns.com,http://www.portfolioprobe.com/blog,http://quantivity.wordpress
.com,http://timelyportfolio.blogspot.com
http://www.theanalysisfactor.com;
http://blogs.sas.com/content/iml;
http://www.johndcook.com/blog;
http://kbroman.wordpress.com;http://www.andrewgelman.com,http://simplystat
istics.org
http://www.cscs.lsa.umich.edu/~crshalizi/weblog;
http://trinkerrstuff.wordpress.com;http://botthoughts.wordpress.com,http:/
/www.r-bloggers.com
http://trinkersstatsstuff.wordpress.com;
http://www.win-vector.com/blog;
http://xianblog.wordpress.com;http://www.andrewgelman.com,http://errorstat
istics.com,http://freakonometrics.blog.free.fr,http://normaldeviate.wordpr
ess.com,http://r-
bloggers.com,http://www.statsblogs.com,http://terrytao.wordpress.com,http:
//gowers.wordpress.com
http://yihui.name/en;
http://observationalepidemiology.blogspot.com;http://www.johndcook.com/blo
g,http://www.andrewgelman.com/blog,http://permut.wordpress.com,http://mrep
idemiology.com,http://educationandstatistics.blogspot.com
http://cooldata.wordpress.com;http://analytics4business.wordpress.com,http
://danallenby.wordpress.com,http://blog.data-
miners.com,http://abbottanalytics.blogspot.com,http://lovestats.wordpress.
com,http://www.datamininglblog.com,http://www.analysisfactor.com,http://www
.statsmakemecry.com,http://statswithcats.wordpress.com,http://timsalimans.
com
http://www.thejuliagroup.com/blog;http://www.andrewgelman.com,http://pleio
tropy.fieldofscience.com,http://lovestats.wordpress.com,http://blogs.sas.c
om/sasdummy,http://simplystatistics.org,http://www.johndcook.com/blog
http://www.flowingdata.com;
http://understandinguncertainty.org;http://www.informationisbeautiful.net,
http://junkcharts.typepad.com,http://tierneylab.blogs.nytimes.com,http://n
ews.bbc.co.uk/1/hi/magazine/7883619.stm,http://blogs.wsj.com/numbersguy,ht
tp://www.andrewgelman.com,http://infosthetics.com
http://scharrheds.blogspot.com;
http://datajournalists.org;
http://brainstat.blogspot.com;http://www.cscs.lsa.umich.edu/~crshalizi/web
log,http://nuit-
blanche.blogspot.com,http://blog.smola.org,http://geomblog.blogspot.com
http://rbaltman.wordpress.com;
http://chrisladroue.com;http://gettinggeneticsdone.blogspot.com,http://www
.flowingdata.com,http://www.informationisbeautiful.net,http://www.r-
bloggers.com,http://www.andrewgelman.com,http://www.johndcook.com/blog,htt
p://statisticsforum.wordpress.com
http://dahuasky.wordpress.com;
http://ivory.idyll.org/blog
http://scienceblogs.com/digitalbio;
http://dirk.eddelbuettel.com;

http://blog.echen.me;
http://hao1990.blogspot.com;
http://heuristically.wordpress.com;http://www.r-bloggers.com
http://www.homolog.us/blogs;http://bcbio.wordpress.com,http://gettinggenet
icsdone.blogspot.com,http://www.r-
bloggers.com,http://www.thegeneticgenealogist.com,http://ongenes.blogspot.
com,http://my.biotechlife.net,http://omicsomics.blogspot.com,http://ongene
tics.blogspot.com,http://bergmanlab.smith.man.ac.uk/?page_id=45,http://hal
danessieve.org,http://flxlexblog.wordpress.com,http://www.jamesandthegiant
corn.com,http://jermdemo.blogspot.com,http://nextgenseq.blogspot.com,http:
//pathogenomics.bham.ac.uk/blog,http://pbeltrao.blogspot.com,http://thegen
omefactory.blogspot.com.au,http://fungalgenomes.org/blog,http://phylogenom
ics.blogspot.com
http://jeromyanglim.blogspot.com;http://www.andrewgelman.com,http://blogs.
mbs.edu/fishing-in-the-bay,http://www.aliquote.org,http://www.daniel-
lemire.com,http://blog.revolutionanalytics.com,http://www.drewconway.com/z
ia,http://www.cerebralmastication.com,http://www.johndcook.com/blog,http:/
/www.johnmyleswhite.com,http://normaldeviate.wordpress.com,http://nsaunder
s.wordpress.com,http://robjhyndman.com,http://www.r-statistics.com
http://www.cnblogs.com/jerrylead;
https://wiki.hpcc.msu.edu/display/~johnj@msu.edu/;
http://www.johnmyleswhite.com;
http://linbaba.wordpress.com;http://almostsure.wordpress.com,http://bit-
player.org,http://djalil.chafai.net/blog,http://xianblog.wordpress.com,htt
p://blogs.williams.edu/Morgan,http://gilkalai.wordpress.com,http://stochas
tix.wordpress.com
http://kevin-
gattaca.blogspot.com;http://phylogenomics.blogspot.com,http://omicsomics.b
logspot.com,http://bcbio.wordpress.com,http://www.clcngs.com,http://hackma
p.blogspot.com
http://ngs-
brescia.blogspot.it;http://omicsomics.blogspot.com,http://phylogenomics.bl
ogspot.com,http://spittoon.23andme.com,http://blog.openhelix.eu,http://ngs
-
expert.com,http://www.thegeneticgenealogist.com,http://seqonomics.blogspot
.it,http://rna-seqblog.com,http://kevin-
gattaca.blogspot.com,http://mygenomix.wordpress.com
http://omicsomics.blogspot.com;http://core-
genomics.blogspot.com,http://www.edgebio.com/blog,http://blog.fejes.ca,htt
p://www.genomesunzipped.org,http://www.genomicslawreport.com,http://www.ma
ssgenomics.org,http://blog.malde.org,http://blogs.forbes.com/matthewherper
,http://kevin-
gattaca.blogspot.com,http://seqanswers.com,http://pathogenomics.bham.ac.uk
/blog,http://www.ngsleaders.org,http://thegenesherpa.blogspot.com,http://d
igitheadslabnotebook.blogspot.com,http://nextgenseq.blogspot.com,http://da
ily-
scala.blogspot.com,http://www.eyeondna.com,http://www.genomeweb.com,http:/
/scienceblogs.com/digitalbio,http://pbeltrao.blogspot.com,http://www.linke
din.com/in/keithrobison,http://bpr3.org/?page_id=56,http://genesearch.word
press.com,http://www.thepersonalgenome.com,http://migration.wordpress.com,
http://scienceblogs.com/evolgen,http://www.geneticsandhealth.com,http://ww
w.pandasthumb.org,http://www.medbioworld.com/postgenomics_blog
http://blog.goldenhelix.com;
http://pairach.com;

http://www.politigenomics.com;http://www.massgenomics.org
http://polylogblog.wordpress.com;http://hunch.net,http://mybiasedcoin.blog
spot.com,http://nuit-blanche.blogspot.com
http://statmethods.wordpress.com;
http://web.ece.rice.edu/richb;
http://simplystatistics.org;
http://statisticsforum.wordpress.com;
http://www.guardian.co.uk/news/datablog;
http://www.excelcharts.com/blog;
http://www.theusrus.de/blog;http://www.andrewgelman.com,http://www.flowing
data.com,http://infosthetics.com,http://junkcharts.typepad.com/junk_charts
,http://www.r-bloggers.com,http://eagereyes.org
http://visualmethods.blogspot.com;http://infosthetics.com,http://visualcom
plexity.com,http://datamining.typepad.com/data_mining,http://eagereyes.org
,http://www.flowingdata.com,http://www.perceptualedge.com/blog,http://junk
charts.typepad.com/junk_charts,http://simplecomplexity.net
http://www.joeparry.com/blog;
http://designnotes.info;
http://www.visualcomplexity.com/vc/blog;
http://www.maxgadney.com;
http://www.coolinfographics.com;http://www.bestinfographics.co.uk,http://c
ommunicationnation.blogspot.com,http://complexdiagrams.com,http://dailyinf
ographic.com,http://datamining.typepad.com/data_mining,http://www.datapoin
ted.net,http://eagereyes.org,http://www.flowingdata.com,http://igraphicsex
plained.blogspot.com,http://www.infographicsblog.com,http://infographicjou
rnal.com,http://www.infographiclove.com,http://infographicsnews.blogspot.c
om,http://infographicpics.com,http://www.infographicsshowcase.com,http://i
nfographicsite.com,http://infographiq.com,http://infosthetics.com,http://j
unkcharts.typepad.com,http://www.datavisualization.fr,http://newsaesthetic
s.tumblr.com,http://newswithnumbers.com,http://www.serialmapper.com,http:/
/simplecomplexity.net,http://update.snd.org,http://socialmediagraphics.pos
terous.com,http://www.styleandflow.com,http://www.swissinfographics.com,ht
tp://www.usefulcharts.com,http://www.visualcomplexity.com/vc,http://visual
oop.com,http://www.visualnews.com/category/infographics,http://visualthink
map.blogspot.com,http://www.vizworld.com,http://infonewt.com,http://www.br
aintapper.com,http://www.chartsmapsdiagrams.com,http://www.columnfivemedia
.com,http://www.briancragin.com,http://www.datadrivenconsulting.com,http:/
/www.infogra.ph,http://tiffanyfarrant.co.uk,http://ffctn.com,http://www.fl
uxvfx.com/shop/infographics,http://www.folioart.co.uk,http://ejfox.com,htt
p://www.funnelinc.com,http://www.johngrimwade.com,http://www.infogr8.com,h
ttp://infographer.ru/en,http://InfographicDesign.org,http://infographiclab
s.com,http://infographicworld.com,http://www.infojocks.com,http://www.info
rmationisbeautiful.net,http://www.InfoMonkeys.com,http://jess3.com,http://
www.kimberleycrofts.com,http://www.joelertola.com/grfx/index.html,http://m
anyeyes.alphaworks.ibm.com/manyeyes,http://www.mikesudal.com,http://www.ne
oformix.com,http://neomam.com,http://nicolasrapp.com,http://nowsourcing.co
m,http://periscopic.com,http://www.pitchinteractive.com,http://ripetungi.c
om,http://sayitvisually.com,http://www.seewhatyoumean.blogspot.com,http://
serialconsign.com,http://www.staubman.com/index.php,http://timeplots.com,h
ttp://two-
n.com,http://vizthink.com,http://www.wallstats.com,http://www.westwood.edu
/programs/school-of-design/visual-
communications,http://www.mikewirthart.com,http://www.theworldasflatland.n
et,http://xplane.com,http://www.97thfloor.com/social-media/infographics

http://www.r-statistics.com;
http://bit-player.org;
http://www.sankey-
diagrams.com;http://chartporn.org,http://eagereyes.org,http://www.flowingd
ata.com,http://www.niceone.org,http://infosthetics.com,http://junkcharts.t
ypepad.com,http://blog.many-
eyes.com,http://mfadiagrams.blogspot.com,http://statisticalgraphics.blog.c
om,http://perceptualedge.com/blog,http://www.visualcomplexity.com/vc,http:
//www.vizthink.com/blog
http://strangemaps.wordpress.com;
http://www.andypope.info/index.htm;
http://thelogcabin.wordpress.com;http://bigcomputing.blogspot.com,http://w
ww.flowingdata.com/,http://www.r-
bloggers.com,http://blog.revolutionanalytics.com
http://hunch.net;http://machine-
learning.blogspot.com,http://www.andrewgelman.com,http://mark.reid.name/ie
m,http://ml.typepad.com,http://blog.computationalcomplexity.org,http://cra
.org/govaffairs/blog,http://geomblog.blogspot.com,http://michaelnielsen.or
g/blog,http://blog.oddhead.com
http://www.mii.ucla.edu/causality;http://normaldeviate.wordpress.com
http://nuit-
blanche.blogspot.it;http://www.flowingdata.com,http://www.sciencebasedmedi
cine.org,http://machinevision4users.blogspot.com,http://terrytao.wordpress
.com,http://www.2physics.com,http://regularize.wordpress.com,https://norma
ldeviate.wordpress.com,http://mathblogging.wordpress.com,http://lkozma.net
/blog,http://bickson.blogspot.com,http://marchonscience.blogspot.com,http:
//fseoane.net/blog,http://karpathy.ca/myblog,http://media.aau.dk/null_spac
e_pursuits,http://dustingmixon.wordpress.com,http://blog.lupi-
software.com,http://blog.rguha.net,http://blog.thingiverse.com,http://high
lyscalable.wordpress.com,http://hackaday.com,http://tm.durusau.net,http://
www.andrewgelman.com,http://www.johndcook.com/blog,http://freakonometrics.
blog.free.fr/index.php,http://xianblog.wordpress.com,http://image-sensors-
world.blogspot.com,http://terahertztechnology.blogspot.com,http://polylogb
log.wordpress.com,http://mrvacuumtube.blogspot.com,http://rjlipton.wordpre
ss.com,http://hunch.net,http://ergodicity.net,http://brainwindows.wordpres
s.com,http://www.decisionsciencenews.com,http://www.arsmathematica.net,htt
p://gazeinteraction.blogspot.com,http://mark.reid.name/iem,http://www.walk
ingrandomly.com,http://compgen.blogspot.com,http://www.kinecthacks.net,htt
p://djalil.chafai.net/blog,http://11011110.livejournal.com,http://shom83.b
logspot.com,http://harvestimaging.com/blog,http://www.computersdontsee.net
,http://scienceinthesands.blogspot.com,http://davidketcheson.info,http://m
ysliceofpizza.blogspot.com,http://laurent-
duval.blogspot.com,http://geomblog.blogspot.com,http://secretsofconsulting
.blogspot.com,http://pixel-
shaker.fr,http://geekmusfir.wordpress.com,http://linbaba.wordpress.com,htt
p://hao1990.blogspot.com,http://nlpers.blogspot.com,http://malkarouri.word
press.com,http://blog.gillerinvestments.com,http://bpchesney.org,http://th
ousandfold.net/cz,http://www.neurevolution.net,http://infostructuralist.wo
rdpress.com,http://quantombone.blogspot.com,http://stupidmatlabhacks.tumbl
r.com,http://lousodrome.net/blog,http://yaroslavvb.blogspot.com,http://dam
arisbsarria.blogspot.com,http://blog.sigfpe.com,http://mirror2image.wordpr
ess.com,http://www.futurepicture.org,http://brainstat.blogspot.com,http://
www.giseliramos.com.br/blog,http://media.aau.dk/CRISSP,http://blog.informa
tiongeometry.org,http://blog.stodden.net,http://espacevide.net/articles,ht

tp://yetaspblog.wordpress.com,http://micheleguieu.blogspot.com,http://marc
iomarim.com/blog,http://people.umass.edu/mduarte/Main/Main.html,http://www
.corrada.com/blog,http://spectralholes.blogspot.com,http://blog.i2pi.com,h
ttp://to-
cs.blog.sohu.com,http://opticalimaging.org/OISblog,http://numberblog.wordp
ress.com,http://electronsandholes.blogspot.com,http://blogperso.univ-
rennes1.fr/arthur.charpentier,http://suksmono.wordpress.com,http://yall1.b
logs.rice.edu,http://hashimotolaboratory.blogspot.com,http://alittleknowle
dge.wordpress.com,http://guiuestc.blogspot.com,http://www.willowgarage.com
/blog,http://cscs.umich.edu/~crshalizi/weblog
http://thestatsblog.wordpress.com;http://blogs.wsj.com/numbersguy
http://www.statschat.org.nz;http://underpoint05.wordpress.com,http://www.b
adscience.net,http://blogstats.wordpress.com,http://chartsnthings.tumblr.c
om,http://citizen-
statistician.org,http://www.datagenetics.com/blog.html,http://blogs.mbs.ed
u/fishing-in-the-
bay,http://fivethirtyeight.blogs.nytimes.com,http://www.flowingdata.com,ht
tp://www.freakonomics.com/blog,http://infosthetics.com,http://junkcharts.t
ypepad.com,http://liesdamnedliesstatistics.com,http://junkcharts.typepad.c
om/numbersruleyourworld,http://simplystatistics.org,http://www.iq.harvard.
edu/blog/sss,http://www.andrewgelman.com,http://statisticsforum.wordpress.
com,http://statswithcats.wordpress.com,http://thestatsblog.wordpress.com,h
ttp://www.straightstatistics.org,http://blogs.abcnews.com/thenumbers,http:
//blogs.wsj.com/numbersguy,http://understandinguncertainty.org/blog
http://davegiles.blogspot.com;
http://romainfrancois.blog.free.fr;http://planetr.stderr.org,http://www.r-
bloggers.com
http://ryouready.wordpress.com;http://www.cerebralmastication.com,http://l
earnr.wordpress.com,http://www.markheckmann.de,http://onertipaday.blogspot
.com,http://www.r-
bloggers.com,http://www.rensenieuwenhuis.nl,http://blog.revolutionanalytic
s.com,http://using-r-
project.blogspot.com,http://worldofrcraft.blogspot.com
http://rdataviz.wordpress.com;
http://societytosupressthecorrelationcoefficient.wordpress.com;http://psyc
hologicalstatistics.blogspot.com,http://seriousstats.wordpress.com
http://bps-
msc.blogspot.com;http://www.andrewgelman.com,http://www.johndcook.com/blog
,http://www.decisionsciencenews.com,http://psychologicalstatistics.blogspo
t.com
http://www.walkingrandomly.com;http://threesixty360.wordpress.com,http://q
chu.wordpress.com,http://crossedstreams.com/wordpress,http://myreckonings.
com/wordpress,http://gottwurfelt.wordpress.com,http://letsplaymath.wordpre
ss.com,http://yaroslavvb.blogspot.com,http://mathnotations.blogspot.com,ht
tp://mathpages.blogspot.com,http://www.mathpuzzle.com,http://physical-
thought.blogspot.com,http://pleasemakeanote.blogspot.com,http://www.mathfi
nance.cn,http://stochastix.wordpress.com,http://scienceandreason.blogspot.
com,http://shuisman.com,http://teachingcollegemath.com,http://www.terminal
ly-
incoherent.com/blog,http://www.johndcook.com/blog,http://www.mathlesstrave
led.com,http://unapologetic.wordpress.com,http://blog.plover.com,http://we
bmaths.wordpress.com,http://wildaboutmath.com
http://www.decisionsciencenews.com;http://messymatters.com,http://www.r-
bloggers.com,http://www.andrewgelman.com/blog

http://permut.wordpress.com;http://www.iq.harvard.edu/blog/sss,http://www.andrewgelman.com,http://codeandculture.wordpress.com,http://www.cscs.umich.edu/~crshalizi/weblog,http://orgtheory.wordpress.com,http://www.overcomingbias.com
http://educationandstatistics.blogspot.com;
http://danallenby.wordpress.com;
http://abbottanalytics.blogspot.com;http://matlabdatamining.blogspot.com,http://timmanns.blogspot.com,http://dataminingblog.com,http://minethatdata.blogspot.com
http://lovestats.wordpress.com;
http://statswithcats.wordpress.com;
http://tierneylab.blogs.nytimes.com;http://www.freakonomics.com/blog
http://blogs.wsj.com/numbersguy;
http://blog.smola.org;
http://geomblog.blogspot.com;http://gilkalai.wordpress.com,http://blog.computationalcomplexity.org,http://lemire.me/blog,http://godplaysdice.blogspot.com,http://mark.reid.name/iem,http://hunch.net,http://nlpers.blogspot.com,http://terrytao.wordpress.com,http://polylogblog.wordpress.com,http://gottwurfelt.wordpress.com
http://almostsure.wordpress.com;http://gottwurfelt.wordpress.com,http://lewko.wordpress.com,http://terrytao.wordpress.com
http://djalil.chafai.net/blog;http://gilkalai.wordpress.com,http://fabricebaudoin.wordpress.com,http://freakonometrics.blog.free.fr,http://linbaba.wordpress.com,http://nuit-blanche.wordpress.com,http://stochastix.wordpress.com,http://gowers.wordpress.com,http://terrytao.wordpress.com
http://blogs.williams.edu/Morgan;
http://gilkalai.wordpress.com;http://weblog.fortnow.com,http://geomblog.blogspot.com,http://blogs.williams.edu/Morgan,http://hunch.net,http://michaelnielsen.org/blog,http://mybiasedcoin.blogspot.com,http://quomodocumque.wordpress.com,http://www.andrewgelman.com,http://terrytao.wordpress.com
http://stochastix.wordpress.com;http://conflate.net/inductio
http://mybiasedcoin.blogspot.com;http://geomblog.blogspot.com,http://geomblog.blogspot.com,http://weblog.fortnow.com
http://chartporn.org;http://www.coolinfographics.com,http://www.flowingdata.com,http://www.informationisbeautiful.net
http://ml.typepad.com;
http://blog.computationalcomplexity.org;http://hunch.net,http://mybiasedcoin.blogspot.com,http://blog.oddhead.com,http://geomblog.blogspot.com,http://polylogblog.wordpress.com,http://terrytao.wordpress.com
http://fseoane.net/blog;
http://karpathy.ca/myblog;
http://chartsnthings.tumblr.com;
http://citizen-statistician.org;http://www.flowingdata.com,http://simplystatistics.org,http://www.statschat.org.nz
http://www.datagenetics.com/blog.html;
http://blogs.mbs.edu/fishing-in-the-bay;
http://liesdamnedliesstatistics.com;http://datamining.typepad.com,http://www.hilarymason.com,http://blog.infochimps.com,http://infosthetics.com,http://www.informationisbeautiful.net,http://www.marketingcharts.com,http://junkcharts.typepad.com/numbersruleyourworld,http://www.statisticsblog.com,http://www.statista.com,http://statosphere.misentropy.com,http://www.comscor

edatamine.com,http://blogs.forrester.com/market_insights,http://thestatsblog.wordpress.com
http://blogs.abcnews.com/thenumbers
http://codeandculture.wordpress.com;http://permut.wordpress.com
http://lewko.wordpress.com;http://terrytao.wordpress.com
http://fabricebaudoin.wordpress.com;http://terrytao.wordpress.com,http://probweb.berkeley.edu,http://gowers.wordpress.com