

**Frequency Analysis on Department Webpages**

To do this analysis I used a program designed to do word counts on web pages, intended to be used for doing indexing (like google does). The program strips out common html tags and punctuation before counting the words. The whole approach is still fairly crude, but it works fairly well.

Note that this is not fully representative of the websites since it is incredibly difficult to do analysis on images, which composed a large part of all the departments' websites.

Here are the results:

Letters	A	B	C	D	E	F	G	H
CSE	9.5919	1.1762	5.7625	5.5898	10.3208	2.3766	2.5656	2.4559
MATH	8.1378	2.9809	3.9245	3.5206	10.0412	2.1331	1.7931	3.4787
Physics	7.4168	2.0247	4.3521	3.6697	11.159	2.2504	1.8266	3.6476
I	J	K	L	M				
6.2023	0.4521	0.70451	3.2044	3.3577				
6.8891	0.296	0.5537	3.803	3.8356				
7.3584	0.2376	0.4313	4.2175	3.0271				

Letters	N	O	P	Q	R	S	T	U
CSE	6.7495	5.1421	3.5569	0.1148	6.5382	8.4588	6.4606	5.2134
Math	7.9367	6.7113	4.2755	0.4112	5.9604	7.8545	7.9967	3.4104
Physics	6.8595	7.2274	3.5648	0.2165	6.2998	7.7801	8.2311	3.4749
V	W	X	Y	Z				
1.1472	1.346	0.1686	1.1682	0.1925				
0.913	1.2497	0.3319	1.4078	0.1521				
1.2109	1.1784	0.4898	1.7023	0.1454				

Total # of letters:	CSE	Math	Phys
	91135	66730	87695
Unique # of words	5243	7824	6569
Entropy:	CSE	Math	Physics
	4.227507	4.251094	4.233528

The interesting part here is all the websites have a low entropy when compared to standard English, which means web traffic may be easier to decrypt. Also, as Professor O'Bryant predicted, the math department has a (slightly) higher entropy than the other two departments.