USING GRAPHS TO BREAK VIGENÈRE CIPHERS

Rick Klima
Department of Mathematical Sciences
Appalachian State University
345 Walker Hall
Boone, North Carolina 28608
klimare@appstate.edu

Neil Sigmon
Department of Mathematics
Radford University
212 Walker Hall
Radford, Virginia 24142
npsigmon@radford.edu

A method for disguising information so that ideally it cannot be understood by anyone but the intended recipient is called a *cipher*. *Cryptanalysis* refers to the process of an unintended recipient of disguised information attempting to remove the disguise and understand the information, and successful cryptanalysis is called *breaking* a cipher.

When a cipher is used to exchange information, the undisguised information is called the *plaintext*, and the disguised information the *ciphertext*. The process of converting from plaintext to ciphertext is called *encryption*. Upon receiving a ciphertext, the recipient must remove the disguise, a process called *decryption*. To be able to effectively encrypt and decrypt messages, correspondents must typically share knowledge of a secret *key*, which is used in applying the cipher. More specifically, the key for a cipher is information usually known only to the originator and intended recipient of a message, which is used by the originator to encrypt the plaintext, and the recipient to decrypt the ciphertext.

Shift Ciphers

For *shift* ciphers, users agree upon an order for the alphabet letters, like for instance the natural order A, B, C, … , Z of letters in our alphabet, and then encrypt each plaintext letter by replacing it with the letter some agreed-upon number of positions to the right in the alphabet, wrapping from the end of the alphabet to the start whenever necessary. For example, for a shift cipher with our alphabet letters in the natural order and in which each plaintext letter is replaced with the letter three positions to the right, the plaintext letter A would be replaced with the letter D, the plaintext letter B with E, C with F, … , W with Z, X with A, Y with B, and Z with C.

Example 1. Consider a shift cipher with our alphabet letters in the natural order and a shift of three positions to the right for encryption. This yields the following plaintext/ciphertext letter pairs.

Plain:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher:	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Using this cipher, the plaintext I CAME, I SAW, I CONQUERED encrypts to the ciphertext LFDPH LVDZL FRQTX HUHG. □

Cryptanalysis of Shift Ciphers

For a message written using our alphabet letters and encrypted with a shift cipher, the ciphertext could result from a maximum of only 25 distinct shifts (assuming a shift of zero positions is not used). Thus, to break the cipher, one could simply try to decrypt the ciphertext assuming each of these 25 possible encryption shifts one at a time, stopping when the correct plaintext is revealed. It would almost certainly be known immediately when the correct plaintext was revealed, since of the results of the various attempts at decryption, it is almost certain that only the letters in the correct plaintext would make sense when strung together. In addition, it may be possible to save a significant amount of time by trying to decrypt just a small portion of the ciphertext, and then decrypting the full ciphertext only after the correct shift is determined.

Example 2. Consider the ciphertext HVSDF CPZSA KWHVG CQWOZ WGAWG HVOHS JSBHI OZZMM CIFIB CIHCT CHVSF DSCDZ SGACB SM, which was formed using a shift cipher with our alphabet letters in the natural order. For each number of possible positions shifted to the right for encryption that could produce this ciphertext, the following shows the result of trying to decrypt the first 10 letters in the ciphertext, starting with a shift of one position, and stopping when plaintext letters that make sense when strung together are obtained.

Shift = 1:	G U R C E B O Y R Z
Shift = 2:	F T Q B D A N X Q Y
Shift = 3:	E S P A C Z M W P X
Shift = 4:	D R O Z B Y L V O W
Shift = 5:	C Q N Y A X K U N V
Shift = 6:	B P M X Z W J T M U
Shift = 7:	A O L W Y V I S L T
Shift = 8:	Z N K V X U H R K S
Shift = 9:	Y M J U W T G Q J R
Shift = 10:	X L I T V S F P I Q
Shift = 11:	W K H S U R E O H P
Shift = 12:	V J G R T Q D N G O
Shift = 13:	U I F Q S P C M F N
Shift = 14:	T H E P R O B L E M

Thus, the number of positions shifted to the right for encryption was almost certainly 14. Trying to decrypt the rest of the ciphertext for this shift yields the full plaintext: THE PROBLEM WITH SOCIALISM IS THAT EVENTUALLY YOU RUN OUT OF OTHER PEOPLE'S MONEY. □

Vigenère Ciphers

Vigenère ciphers use sequences of shift ciphers, whose use is facilitated through a rectangular array of letters referred to as the *Vigenère square*, shown in Table 1.

	(Plaintext Letter)																								
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Table 1. The Vigenère square

The top row of the Vigenère square consists of the letters A through Z, in order from left to right, representing plaintext letters. The leftmost column of the square also consists of the letters A through Z, in order from top to bottom, representing key letters. With these plaintext letters viewed as column labels and key letters viewed as row labels, the letters in the inner part of the square represent ciphertext letters that correspond to pairs of plaintext and key letters, with the ciphertext letter that corresponds to a particular pair of plaintext and key letters being the letter in the inner part of the square where the column labeled with the plaintext letter intersects the row labeled with the key letter.

Vigenère ciphers require the originator and intended recipient of a message to agree upon one or more keywords. Encryption is then done using the Vigenère square, with the key letters determined by repeating the letters in the keyword(s) as many times as necessary until the total number of key letters matches the total number of plaintext letters.

Example 3. Consider a Vigenère cipher with the keyword TRIXIE. Using this cipher, the plaintext HAVING A PET CAN MAKE YOU HAPPY encrypts as follows.

Plain:	H	A	V	I	N	G	A	P	E	T	C	A	N	M	A	K	E	Y	O	U	H	A	P	P	Y
Key:	T	R	I	X	I	E	T	R	I	X	I	E	T	R	I	X	I	E	T	R	I	X	I	E	T
Cipher:	A	R	D	F	V	K	T	G	M	Q	K	E	G	D	I	H	M	C	H	L	P	X	X	T	R

Thus, the ciphertext is ARDFV KTGMQ KEGDI HMCHL PXXTR. □

Cryptanalysis of Vigenère Ciphers

Note that in Example 3, every sixth plaintext letter is encrypted with the same keyword letter. In addition, due to the shifting pattern of the Vigenère square, each keyword letter in a Vigenère cipher dictates a shift cipher to be applied to any plaintext letter designated to be encrypted with it. For example, for the correspondences A = 0, B = 1, C = 2, … , Z = 25, since T = 19, then any plaintext letter encrypted with the keyword letter T is equivalently encrypted using a shift cipher with our alphabet letters in the natural order and a shift of 19 positions to the right for encryption. Thus, a Vigenère cipher is simply a sequence of shift ciphers, one for each of the letters in the keyword, and the keyword letters themselves can be determined by breaking these shift ciphers.

To break these shift ciphers, we begin by separating a ciphertext into groups in which the letters have all been encrypted with a common keyword letter, and thus a common shift cipher. For instance, in Example 3, if we knew that the length of the keyword was six, we would know that each of the ciphertext letters A, T, G, H, and R resulted from a common shift cipher. Similarly, we would know that each of the ciphertext letters R, G, D, and L resulted from a common shift cipher. These groups of letters are called *cosets*. A full collection of cosets for this ciphertext is ATGHR, RGDL, DMIP, FQHX, VKMX, and KECT.

Each of these cosets contains ciphertext letters that all result from a common shift cipher. Breaking these shift ciphers within a Vigenère cipher is generally more complicated than breaking full shift ciphers, though, since even when decrypted correctly, the plaintext letters that correspond to the letters in a coset should not form sensible English when strung together. Fortunately, several processes for breaking these shift ciphers within a Vigenère cipher have been developed and refined. We will discuss one such process.

We will start by demonstrating a graphical method that involves plotting frequencies to find the length of the keyword for a Vigenère cipher. To begin, consider the frequencies with which the 26 letters in our alphabet occur in ordinary English, shown in Table 2.

Letter	Frequency	Letter	Frequency
A	0.0817	N	0.0675
B	0.0149	O	0.0751
C	0.0278	P	0.0193
D	0.0425	Q	0.0010
E	0.1270	R	0.0599
F	0.0223	S	0.0633
G	0.0202	T	0.0906
H	0.0609	U	0.0276
I	0.0697	V	0.0098
J	0.0015	W	0.0236
K	0.0077	X	0.0015
L	0.0403	Y	0.0197
M	0.0241	Z	0.0007

Table 2. Letter frequencies in ordinary English

The *signature* of ordinary English is a graph of these frequencies, plotted from smallest to largest, with each pair of consecutive points connected by a straight line. The signature of ordinary English is shown in Figure 1. This graph starts on the left even with the smallest frequency 0.0007 in Table 2, and ends on the right even with the largest 0.1270.

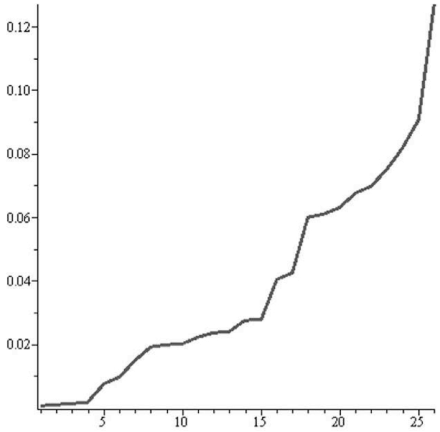

Figure 1. Signature of ordinary English

Similarly, the signature of a sample of text is a graph of the frequencies with which the letters occur in the sample. Two facts are important when considering the signature of a sample. First, in a sample it is possible that some letters will not occur. The frequencies of these letters would then be 0. Second, since the sum of all frequencies, whether for the full English language or just a sample, must be 1, to compensate for the existence of letters with frequency 0 in a sample, it is likely that some letters will have higher frequencies than in ordinary English. Thus, the signature of a sample should in general be lower than the signature of ordinary English at the start of the graph (on the left), and higher at the end. To illustrate this, consider the plaintext `WE WANT TO LOOK AT THE SIGNATURE OF THE SAMPLE OF A MESSAGE`. The frequencies with which the 26 letters occur in this plaintext are shown in Table 3.

Letter	Frequency	Letter	Frequency
A	0.1277	N	0.0426
B	0	O	0.1064
C	0	P	0.0213
D	0	Q	0
E	0.1489	R	0.0213
F	0.0426	S	0.0851
G	0.0426	T	0.1277
H	0.0426	U	0.0213
I	0.0213	V	0
J	0	W	0.0426
K	0.0213	X	0
L	0.0426	Y	0
M	0.0426	Z	0

Table 3. Letter frequencies in a sample plaintext

Figure 2 shows the signature of this plaintext (the thinner segments) along with the signature of ordinary English (the thicker segments). Notice that the signature of the plaintext is indeed lower than the signature of ordinary English at the start of the graph due to the frequencies of 0 in the plaintext, and higher at the end in compensation.

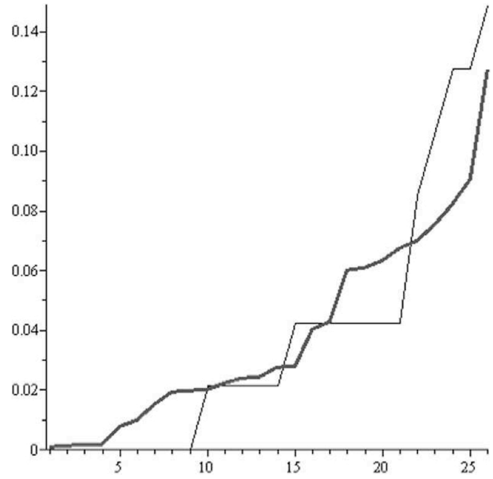

Figure 2. Signature of ordinary English and a sample

Another important fact when considering the signature of a sample is that since shift ciphers preserve letter frequencies from plaintexts into ciphertexts, then for a ciphertext formed using a shift cipher, the signature of the ciphertext will be identical to the signature of the plaintext. To illustrate this, consider again the plaintext `WE WANT TO LOOK AT THE SIGNATURE OF THE SAMPLE OF A MESSAGE`, encrypted using a shift cipher with our alphabet letters in the natural order and a shift of 19 positions to the right for encryption, for which the resulting ciphertext is `PXPTG MMHEH HDTMM AXLBZ GTMNK XHYMA XLTFI EXHYT FXLLT ZX`. The frequencies with which the 26 letters in our alphabet occur in this ciphertext are shown in Table 4.

Letter	Frequency	Letter	Frequency
A	0.0426	N	0.0213
B	0.0213	O	0
C	0	P	0.0426
D	0.0213	Q	0
E	0.0426	R	0
F	0.0426	S	0
G	0.0426	T	0.1277
H	0.1064	U	0
I	0.0213	V	0
J	0	W	0
K	0.0213	X	0.1489
L	0.0851	Y	0.0426
M	0.1277	Z	0.0426

Table 4. Letter frequencies in a sample ciphertext

Note that Tables 3 and 4 contain identical frequencies, but associated with different letters. Thus, the signature of this ciphertext would be identical to the signature of the plaintext. That is, Figure 2, which we noted previously shows the signatures of the plaintext and ordinary English, also shows the signatures of the ciphertext and ordinary English.

Now, recall that each keyword letter dictates a shift cipher to be applied to any plaintext letter designated to be encrypted with it. Recall also that the ciphertext letters encrypted with a common shift cipher form a coset. If a ciphertext is separated into the correct number of cosets, then the signatures of these cosets should start lower and end higher when compared with the signature of ordinary English. On the other hand, if the ciphertext is separated into an incorrect number of cosets, then the signatures of these cosets should fail to exhibit this behavior. Looking for this behavior can help identify the likely correct number of cosets, which would then be the likely length of the keyword.

Example 4. Consider the ciphertext `UZRZE GNJEN VLISE XRHLY PYEGT ESBJH JCSBP TGDYF XXBHE EIFTC CHVRK PNHWX PCTUQ TGDJH TBIPR FEMJC NHVTC FSAII IFNRE GSALH XHWZW RZXGT TVWGD HTEYX ISAGQ TCJPR SIAPT UMGZA LHXHH SOHPW CZLBR ZTCBR GHCDI QIKTO AAEFT OPYEG TENRA IALNR XLPCE PYKGP NGPRQ PIAKW XDCBZ XGPDN RWXEI FZXGJ LVOXA JTUEM BLNLQ HGPWV PEQPI AXATY ENVYJ EUEI`, which was formed using a Vigenère cipher. Figure 3 shows the signatures of the cosets (the thinner segments) and ordinary English (the thicker segments) when this ciphertext is separated into four, five, six, and seven cosets.

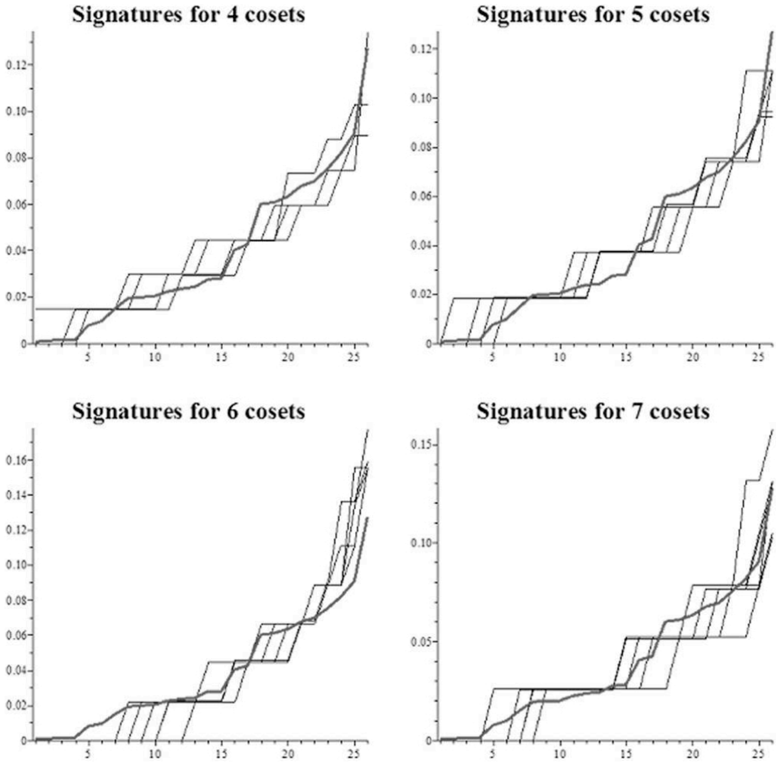

Figure 3. Signatures of ordinary English and cosets

The number of cosets whose signatures best exhibit the behavior of starting lower and ending higher than the signature of ordinary English is six. Thus, the likely length of the keyword for the cipher is six. □

Once the number of letters in the keyword for a Vigenère cipher is known, these letters must be determined. To do this, consider again the frequencies with which the 26 letters in our alphabet occur in ordinary English, shown in Table 2. The scrawl of ordinary English is like the signature, except the frequencies are plotted in alphabetical order. The scrawl of ordinary English is shown in Figure 4. This graph starts on the left even with the first frequency 0.0817 in Table 2, and ends on the right even with the last 0.0007.

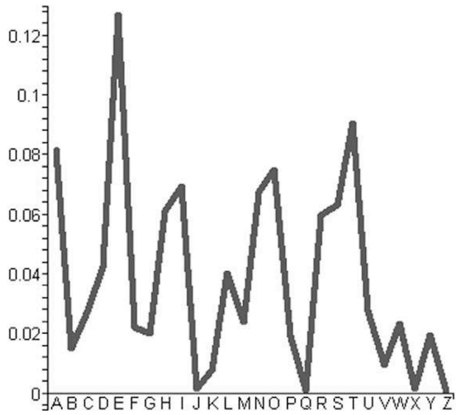

Figure 4. Scrawl of ordinary English

Similarly, the scrawl of a sample of text is like the signature, but with the frequencies plotted in alphabetical order. For a plaintext or for a ciphertext formed using a shift cipher, even for a relatively small sample, the scrawl of the sample should have a similar appearance to the scrawl of ordinary English, with roughly the same peaks and dips. However, for a coset in a ciphertext formed using a Vigenère cipher, recall that the letters all result from a common shift cipher. So while the scrawl of a coset should have roughly the same peaks and dips as the scrawl of ordinary English, these peaks and dips should all be shifted to the right some number of positions, wrapping from the right edge of the graph to the left when necessary, with the number of positions shifted corresponding to the keyword letter that dictated the shift.

More precisely, with the correspondences `A` = 0, `B` = 1, `C` = 2, … , `Z` = 25, for a particular keyword letter, all ciphertext letters encrypted with the keyword letter will be shifted to the right the corresponding number of positions. For the coset containing these ciphertext letters, if the scrawl of the coset were shifted to the left the same number of positions, wrapping from the left edge of the graph to the right when necessary, the scrawl should be as closely aligned as possible with the scrawl of ordinary English. Thus, shifting the scrawl of a coset, looking for where it aligns as closely as possible with the scrawl of ordinary English, can help identify the keyword letter that produced the shift. Doing this for each coset should allow the entire keyword to be determined one letter at a time.

Example 5. Consider again the ciphertext in Example 4, which was formed using a Vigenère cipher, and for which in Example 4 we found that the likely length of the keyword for the cipher is six. To determine the keyword letters, we begin by separating the ciphertext into the following six cosets.

```
Coset 1:  UNILTJGBTKPGIJCISWGDICAZHCTCTTTAPGQXGXGABGQTJ
Coset 2:  ZJSYECDHCPCDPCFFAZTHSJPASZCDOOELCPPDPEJJLPPYE
Coset 3:  REEPSSYECNTJRNSNLWTTAPTLOLBIAPNNENICDILTNWIEU
Coset 4:  ZNXYBBFEHHUHFHARHRVEGRUHHBRQAYRRPGABNFVULVANE
Coset 5:  EVREJPXIVWQTEVIEXZWYQSMXPRGIEEAXYPKZRZOEQPXVI
Coset 6:  GLHGHTXFRXTBMTIGHXGXTIGHWZHKFGILKRWXWXXMHEAY
```

Figure 5 shows, for each of these cosets individually, the scrawls of the coset (the thinner segments) and ordinary English (the thicker segments) for two shifts, one being the correct shift, and the other chosen randomly.

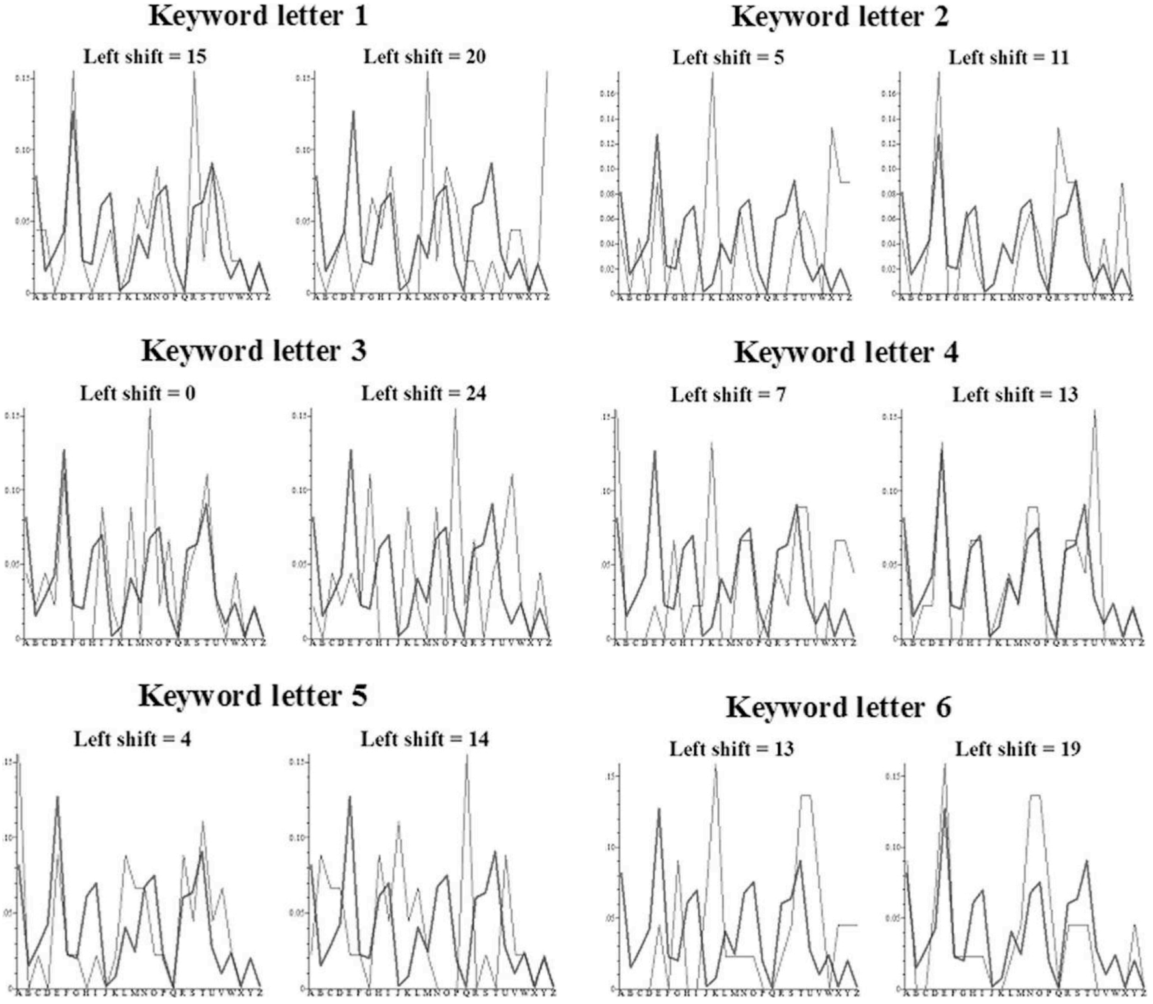

Figure 5. Scrawls of ordinary English and cosets

The shifts for which the scrawls of the cosets more closely align with the scrawls of ordinary English are, in order, 15, 11, 0, 13, 4, and 19 positions. With the correspondences A = 0, B = 1, C = 2, … , Z = 25, these shifts give the keyword PLANET. With this keyword, the ciphertext decrypts to: FOR MANY YEARS, THE KNOWN PLANETS OF OUR SOLAR SYSTEM WERE MERCURY, VENUS, EARTH, MARS, JUPITER, SATURN, URANUS, NEPTUNE, AND PLUTO. HOWEVER, IT IS NOW TRUE THAT MANY PEOPLE THINK PLUTO SHOULD NO LONGER BE CONSIDERED A NAMED PLANET. NEW PLANETS ARE CURRENTLY BEING DISCOVERED, AND IT IS VERY LIKELY THAT MANY MORE WILL BE IN THE NEAR FUTURE. □

Choosing the shifts that lead to the correct keyword letters in Example 5 was made much easier by the fact that only two alignments had to be considered for each letter, as opposed to the 26 that would have to be considered for each letter in actual practice. This method for breaking Vigenère ciphers is obviously reliant upon the production of many carefully constructed graphs. While it would be possible to draw these graphs by hand, since signatures and scrawls are just plotted points connected by straight lines, it might not be practical to do so. Fortunately, in modern society, technology exists that facilitates the production of these graphs.

Through the graphing functions readily available within computer algebra systems such as Maple, it is relatively easy to use this process for cryptanalysis of Vigenère ciphers. Due to space limitations, it is not possible for us to demonstrate this here. However, the Maplet VigenereCipherBreaker.maplet, which we wrote to use this process for cryptanalysis of Vigenère ciphers, and which we demonstrated at ICTCM 2012, can be downloaded from [2]. The PowerPoint file we used to accompany this demonstration can be downloaded from [4]. This process for cryptanalysis of Vigenère ciphers is also demonstrated in [1] and [3], and our Maplet is also described and demonstrated in [3].

Conclusion

We have shown how graphs can be used to assist in the cryptanalysis of Vigenère ciphers, which would be appropriate as a part of any general education mathematics course.

References

[1] T. Barr. *Invitation to Cryptology*. Prentice Hall, Upper Saddle River, NJ, 2002.

[2] R. Klima and N. Sigmon. A Maplet for cryptanalysis of Vigenère ciphers. Available at http://www.appstate.edu/~klimare/VigenereCipherBreaker.maplet.

[3] R. Klima and N. Sigmon. *Cryptology, Classical and Modern, with Maplets.* Taylor & Francis/CRC Press, Boca Raton, FL, 2012.

[4] R. Klima and N. Sigmon. PowerPoint file from presentation at ICTCM 2012. Available at http://www.appstate.edu/~klimare/VigenereCipherBreakerPresentation.ppt.