5. A Stemma of Mark in Family 13 using Probability Structure Analysis

G.P. Farthing

This article offers a statement of the basis of Probability Structure Analysis, as I have now developed this conjecture beyond my earlier publications.[1] It also offers my early analysis of the Gospel of Mark in Family 13. Many complications have been omitted since a full treatment would require the scope of a book to examine them. It is my intention to publish in due course a complete analysis of Family 13, at least in Mark, showing what can be discovered by this method about the complex history of these manuscripts and the Family's relationship with the earliest text and the Byzantine text.

---

[1] G.P. Farthing, 'Detailed Textual Stemmata by means of Probability Theory', *Actes du Quatrième Colloque international Bible et informatique: matériel et matière: l'impact de l'informatique sur les études bibliques = Proceedings of the Fourth International Colloquium on the Bible and Computer: Desk and Discipline: The Impact of Computers on Bible Studies: Amsterdam, 15–18 August 1994*, ed. Association Internationale Bible Et Informatique and Vrije Universiteit Amsterdam, Collection DEBORA 8 (Paris: Champion, 1995), pp. 214–222 and G.P. Farthing, 'Using Probability Theory as a Key to Unlock Textual History', *Studies in the Early Text of the Gospels and Acts: The Papers of the First Birmingham Colloquium on the Textual Criticism of the New Testament*, ed. David G.K. Taylor, TS (III) 1 (Birmingham: University of Birmingham Press, 1999; repr. Piscataway, NJ: Gorgias, 2013), pp. 110–134.

In 1942, Kirsopp and Silva Lake gave Figure 1 for Family 13 in Mark, but they added: 'This diagram is concerned solely with the relation of the manuscripts to each other, without consideration for corruption from other texts. To it, to understand the matter fully, must be added influence by the Byzantine text on y, on c and on codex 124, as well as a certain amount of reinfusion of Caesarean readings in 124'.[2]

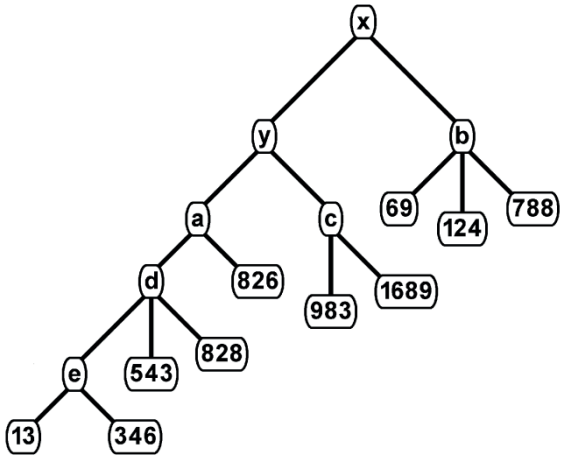

Figure 1. The Lakes' Stemma

Understanding the Lakes' stemma is straightforward. In the Lakes' stemma the lines—also called branches or stems below—each relate to a specific copying event, or groups of copyings, which each yield a specific text combination, so that their stemma, with fifteen stems, involves fifteen real text combinations: ten where one text stands alone and five where several texts stand against the rest. For instance, stem (e) to (d) relates to GA 13, 346 :: 69, 124, 543, 788, 826, 828, 983, 1689; stem (e) to GA 13 relates to GA 13 :: 69, 124, 346, 543, 788, 826, 828, 983, 1689 and stem (y) to (x) to (b), counted as a single stem, relates to GA 13, 346, 543, 826, 828, 983, 1689 :: 69, 124, 788. In other words, there is a one-to-one

[2] Kirsopp Lake and Silva Lake, *Family 13 (The Ferrar Group), Mark with a collation of Codex 28 of the Gospels,* SD 11 (London: Christophers, 1941), p. 42.

correspondence between a specific stem and a specific text combination. However, for ten real texts there are not fifteen but 511 possible text combinations. In fact, over a hundred actual text combinations come from a comparison of the texts of these ten manuscripts, even if only those with two forms of the text are considered. Furthermore, in the fifteen combinations consistent with the Lakes' stemma the texts are divided into two neat groups. For instance, to cite one example of many, in Mark 4:16 there is the simple variation: GA 13, 346, 543, 826, 828: εὐθύς || 69, 124, 788, 983, 1689: εὐθέως. Finding the various manuscripts on the Lakes' stemma it is clear that this variation would most simply occur by a change on the stem (y) to (a).

However, most variations are not so neat. For instance, in Mark 4:32 there is the variation: GA 13, 346, 543, 788, 826: πετεινά || 69, 124, 828, 983, 1689: τὰ πετεινά. The variation is neither significant nor the combination of manuscripts frequent, but it is typical of many variations found in Mark and illustrates the point well. If this combination is plotted graphically on the stemma as in Figure 2, at least four fragments result.

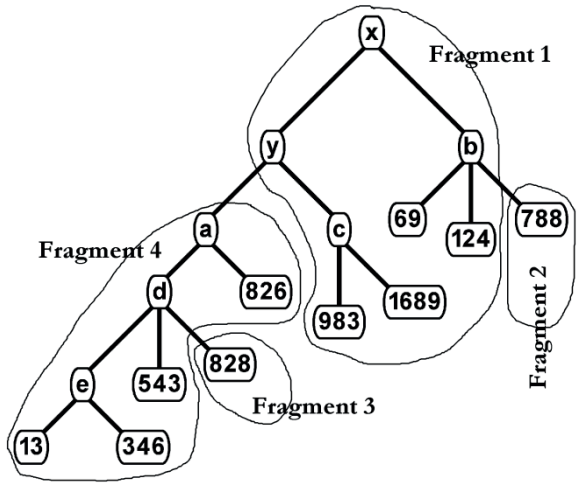

Figure 2

Considering the history of this variation and assuming the earliest text to be found in (x) in Fragment 1, identical changes occur on stems (b) to GA 788 and (y) to (a) because Fragment 2 has the

same texts as Fragment 4. But how does Fragment 3 get its text, since it must agree with Fragment 1? The only way is by the change on stem (y) to (a) being reversed on the stem (d) to GA 828; that is by the scribe of GA 828 writing the text of (x) at this point. These illustrate the processes that make assembling a credible stemma from text combinations so difficult: identical changes do occur and sometimes a scribe, for whatever reason, returns a point in the text to an earlier form. It is important to realise that in all but the fifteen combinations found in collating the texts of these ten manuscripts which conform to the given stemma, one or both sides of the text combination must be fragmented to fit the stemma. Put another way, in most combinations chance links are present due to the same change occurring coincidentally or due to readings being coincidentally reintroduced. For the textual critic attempting to find the underlying history of the manuscripts which can then be represented as a stemma, these occurrences are great difficulties. Most combinations do not fit any simple stemma, and if complex stemmata are attempted where does one stop in including more and more text combinations? These chance agreements are also what makes any attempt at describing the textual relationships based on their similarity very difficult.

Probability Structure Analysis considers a great number of the variations discoverable by collating real manuscript texts. The significance of most of these combinations is fragmentary, as in the example above, but, if these fragments of meaning are combined carefully, a credible underlying history or stemma can be discovered. Probability Structure Analysis seeks to investigate, account for, and represent the more elusive complications alluded to by the Lakes by coordinating these many fragmentary textual relationships. Probability Structure Analysis does not work with the frequencies of the real text combinations directly but seeks to create a model whose text combinations parallel the real text combinations as closely as possible. A good match suggests that the Probability Structure model stemma is a ‘good enough’ representation of the underlying history or stemma of the real manuscripts.

BASIC CONJECTURES OF THE PROBABILITY STRUCTURE MODEL

The following conjectures are the basis on which I have built the models used in this analysis and believe they have been supported by finding useful results so far.

1. That all the texts in the model consist of a finite number of possible points of variation.
2. That each possible point of variation can have two, and only two, forms.
3. That each copying process is characterised by a specific probability of each possible point of variation changing.

The following simple example demonstrates how Probability Structural Analysis works. Let us consider a symbolic text represented by twenty-six possible points of variation each identified by a lower-case letter (Figure 3). Each possible point of variation also has one and only one other form or state, identified by an upper-case letter:

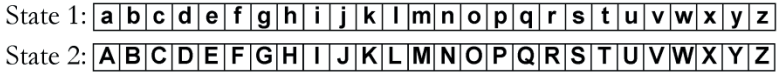

Figure 3

We take a manuscript A with a text where every possible point of variation is in the lower-case form:

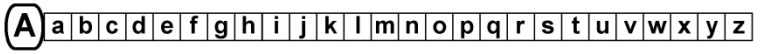

Figure 4

B is a copy of A made with a 40% probability of each possible point of variation changing. 40% of 26 is 10.4, but the result must be a whole number and the nearest whole number is 10. The specific points of variation which change in any scenario are of course random:

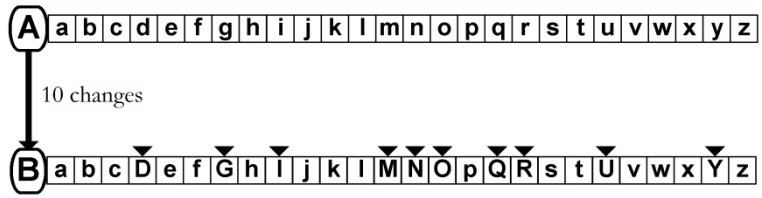

Figure 5

A second copy C is then made independently with a 30% probability of each possible point of variation changing. 30% of 26 is 7.8. The nearest whole number is eight changes. Again the 'choice' of which points of variation change is random:

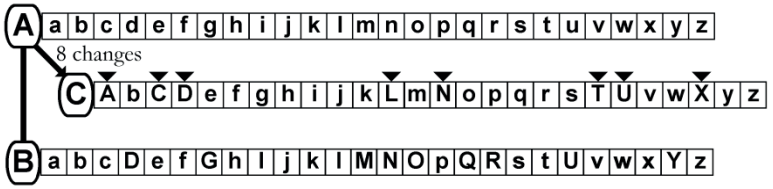

Figure 6

The crucial and fundamental thing to notice is that the changes forming B and the changes forming C coincide three times *by chance* at D, N and U, despite sharing no genealogical relationship:

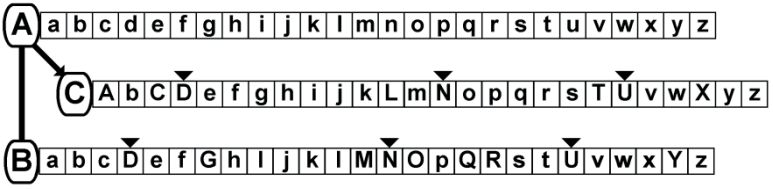

Figure 7

With Probability Structure Analysis, regardless of how few or how many texts are represented, every possible text combination will be assigned a probable frequency, though in some cases the figure may be insignificant. The strength of Probability Structure Analysis is its ability to give frequencies to combinations formed by chance duplications and chance reversals of changes which so bedevil our

understanding of the real texts. The result of this modelling is that text combinations that do not fit the history of the textual copying will always arise; their presence is inevitable. This, I claim, is the aspect of Probability Structure Analysis that correctly mirrors real life; the mass of text combinations that do not conveniently fit any suggested stemma of real texts are not weird aberrations but inevitable consequences of the real copying process. Thus, if we are to model real copying, we need an analysis which models this chaotic aspect as Probability Structure Analysis does.

STEMMATA WITH CONFLATE TEXTS

A further issue that Probability Structure Analysis can deal with is the case where a text has more than one parent. This can arise in various ways. A scribe may find a part of the source is missing.[3] The scribe will then find a second copy to compensate for what is missing. Or a scribe who memorised the text may occasionally 'correct' the text as the copy is made. These circumstances can be called mixture, conflation or contamination. Probability Structure Analysis deals with this circumstance by proportioning two or more stems that feed the conflate text. Clarifying the data is complex but the presentation is quite straightforward (Figure 8). Here C is partly derived from A and partly from B. Dotted stems are given proportions which must add up to exactly one. The full stem is given a probability which converts to a mean number of changes which occur in the copying process. The assumed text X is a convenient fiction to assist calculation.

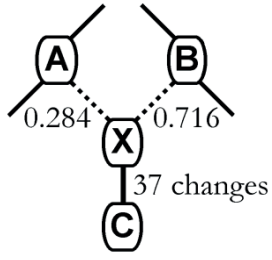

Figure 8

[3]As we will see in 'A Practical Example' below, several manuscripts in this study have small sections of text missing.

THE BASIC MATHEMATICS OF PROBABILITY STRUCTURE ANALYSIS[4]

Let us suppose a set of four manuscripts of a document, called A, B, C and D. These manuscripts are collated and the following number of times each grouping is found is shown in Table 1.

A :: B, C, D	3 times
A, B :: C, D	30 times
A, C :: B, D	15 times
A, D :: B, C	17 times
A, B, C :: D	345 times
A, B, D :: C	260 times
A, C, D :: B	160 times

Table 1. Groupings in four manuscripts

A stemma is suggested which seeks to explain the relationships of these texts (Figure 9).[5]

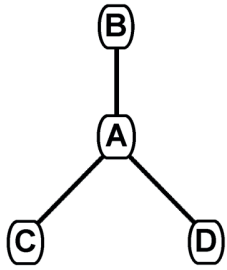

Figure 9

[4] The books from which I learned this mathematics are now well out of print. If the reader wishes to pursue the ideas the internet is full of references (but for that reason is a bit of a maze). I suggest a start at https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests which explains various uses for the chi squared test and helpfully gives a table converting the error figure and degrees of freedom to probabilities in Appendix Table C.pdf. The detail in this table is poor but will give a start for further searches.

[5] A stemma can be constructed manually on inspection of the data, especially if one is building on an existing stemma, or a program can start from scratch and test all possible simple stemmata to find the one offering the lowest error figure. Both methods are used in manipulating the real data in this article.

Giving the computer program the real text combination figures and the suggested stemma, it offers the stem lengths and N figure in Table 2 as the optimum in order most accurately to model the relationships of the real texts based on the group agreement figures. The N figure is the notional number of possible points of variation in the text. The length of the stems given in Table 2 have added to them the probabilities of the text changing on that stem and the complementary probability of the text not changing on that stem. The length of each stem is the mean number of changes expected on that stem, given the probability of change on that stem and the N value: it is the centre of a range and can sensibly be shown with a decimal fraction, whereas the real frequencies must be whole numbers. The letter P is used for the probability of change and Q for the probability of no change. Each individual P plus Q must add to one.

Stem	Length	P of change	Q of no change
A-B	$L_{AB} = 194.27$	$P_{AB} = 0.05087$	$Q_{AB} = 0.94913$
A-C	$L_{AC} = 307.90$	$P_{AC} = 0.08062$	$Q_{AC} = 0.91938$
A-D	$L_{AD} = 392.77$	$P_{AD} = 0.10285$	$Q_{AD} = 0.89715$
The N figure: 3819			

Table 2. Optimum calculations of sample.[6]

The stem lengths can be found by multiplying the appropriate probability of change and the N figure; the probabilities of change can be found by dividing the appropriate stem length by the N figure. The N figure is needed by the computer for its calculations but represents nothing objective in describing the texts of the manuscripts. For this reason, the frequency of the combination where all texts agree is taken to be uncountable and ignored.

Given this stemma, the implied combination frequencies can be calculated. For instance, A and B standing against C and D (A, B :: C, D) occurs where there is no change on the stem A-B (since these texts agree) but there are changes on stems A-C and A-D (since the texts at each end of the stems disagree). The probabilities of change and no change are given in Table 2, as is the number of points of possible variation. The frequency is:

[6] The figures given in this article are rounded to give a neat presentation and may not match exactly in the last significant figure if mathematical operations are performed on them.

Formula 1 $F(A, B :: C,D) = N \times Q_{AB} \times P_{AC} \times P_{AD}$
$= 3819 \times 0.94913 \times 0.08062 \times 0.10285$
$= 30.055$

I have then used a χ^2 (chi squared) calculation to determine the error figure arising from a comparison of the given value (30) and the modelled value (30.055). The general formula is:

Formula 2 $\chi^2 = \frac{(R-M)^2}{R}$

Where R is the frequency derived from the real texts and M from the model. Thus, for F (A, B :: C, D) the error is:

Formula 3 $E(A, B :: C, D) = (30 - 30.055)^2 / 3 = 0.00010$

The other error figures can be calculated in the same way, except for E (A :: B, C, D); here the frequency of three is too small to be used. χ^2 calculations need a figure of at least five to be valid. The solution is to group the 3 for A :: B, C, D with the next smallest figure, 15, which is for A, C :: B, D.[7] This method of grouping, or consolidation, works well for small numbers of figures but gives difficulties for larger numbers of figures; large groupings can conceal large errors. While the χ^2 method is satisfactory for the purposes of demonstration of the method's potential, seeking a more reliable method for larger stemmata is an important next step of my research. The χ^2 method is not in any way intrinsic to Probability Structure Analysis. The full set of results is shown in Table 3.

Combination	Real	Calculated	Error
A :: B, C, D	3	*consolidated with A, C :: B, D*	
A, B :: C, D	30	30.055	0.00010
A, C :: B, D	15 + 3	19.980	0.21779
A, D :: B, C	17	14.052	0.51118
A, B, C :: D	345	342.732	0.01491
A, B, D :: C	260	262.186	0.01838
A, C, D :: B	160	160.239	0.00036
Sum			**0.76271**

Table 3. Error figures for sample

[7] If there are several frequencies of less than five, but which together equal or exceed five, they are treated as a single independent grouping, which will give its own error figure.

This error figure can be converted to a probability that the stemma, with its attributes, is a good explanation of the relationships of the real manuscript texts. However, one more value is needed for the calculations, that is the 'degrees of freedom'. In our example we are trying to account for the relationships of our four manuscript texts based on seven numbers: the frequencies of the various group agreements (ignoring where all texts agree). It is self-evident that if we use a mathematical system which uses seven or more numerical values it would be possible, without fail, to show a convincing comparison. Mathematically, if seven variables are used to account for seven given figures, the 'degrees of freedom' is zero and the attempt pointless. For a meaningful system, fewer values must be used than those to be matched. In our calculations we have six values to account for (there are seven frequencies, but one is consolidated with another, leaving six), and use three stem lengths plus an N value, that is four values, giving a 'degree of freedom' of six minus four which equals two. Given the error figure and the degrees of freedom, standard tables, or a very complicated formula, will give a probability that the calculated figures, with that particular stemma, offer a reasonable explanation of the relationship of the four manuscript texts. The result here is 68.3%. The example's data was, of course, manufactured to give a high probability.

The method of calculating the frequency of combinations where a text has been assumed in order to make sense of the stemma can be demonstrated from Table 3. If we wished to list the combinations of those texts, taking account only of B, C and D, the numbers become clear by listing the seven combinations in Table 3 but omitting the unwanted A, as in Table 4.

Combination	Frequency
:: B, C, D	3
B :: C, D	30
C :: B, D	15
D :: B, C	17
B, C :: D	345
B, D :: C	260
C, D :: B	160

Table 4. Combinations without A

The first combination becomes B, C, D :: where all texts agree, and is ignored. Each of the other useful combinations now occur twice: the second is the same as the seventh, the third the same as the sixth, the fourth the same as the fifth. By adding in pairs, we obtain the following results:

B :: C, D	30	+	C, D :: B	160	= 190
C :: B, D	15	+	B, D :: C	260	= 275
D :: B, C	17	+	B, C :: D	345	= 362

Table 5. Results of combinations without A

The general rule where the stemma has an assumed text is simple: count the frequency, as indicated in the stemma, with the assumed text agreeing with the texts on the left of the combination and add this to the frequency with the assumed text agreeing with the right side. Where there is one assumed text two frequencies will be added; if there are two assumed texts it will be four frequencies; if three texts, it will be eight, and so on, the number being two to the power of the number of assumed texts.

A PRACTICAL EXAMPLE

Given all this, can so simple a system usefully model the relationships of actual texts? This example relates to the same group of texts as the Lakes studied, with the exception of GA 1689. When I began my studies, GA 1689 was considered 'lost'. For instance, Jacob Geerlings in his collation of Family 13 in Luke says: 'The variant readings of 1689 have been extracted from Soden's apparatus and until this manuscript is rediscovered, his apparatus is unfortunately the only source of information about the text of this manuscript'.[8] While this manuscript has now been rediscovered in Prague and is present in the NTVMR, this information came to me too late for inclusion in this article.

To answer the question, I will construct a stemma based on the full transcriptions of the nine texts of Family 13 in Mark, excluding GA 1689, found in the NTVMR. Sadly, three

[8] J. Geerlings, *Family 13 (The Ferrar Group). The Text According to Luke*, SD 20 (Salt Lake City: University of Utah Press, 1961), p. 1.

manuscripts have folios missing: GA 13 is missing 1:21 to 1:45a; GA 543 is missing 8:4b to 8:28a; 826 is missing 12:3b to 12:19a. Since the mathematics requires all texts to be present in each variation examined these sections of Mark, approximately 10% of the text, are not considered. Further, for this study, I have used only variations which have two forms of text and omitted variations consisting wholly of the presence or absence of an *iota* subscript or a final *nu*. Thus, while the transcription of the texts is the work made available by the INTF, the collation and counting of variations is entirely my responsibility. My analysis program, using these data, offers the stemma in Figure 10 as the best simple stemma—a stemma lacking any conflate text—for Family 13 in Mark, with the Lakes' stemma repeated as Figure 11 for comparison.

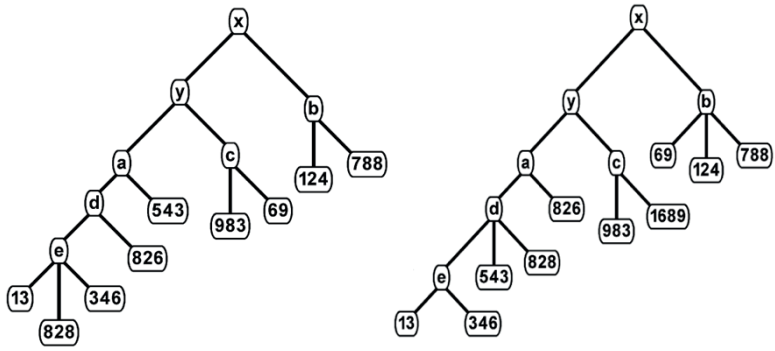

Figure 10. Stemma by Farthing (left)
Figure 11. The Lakes' stemma (right)

Figure 10 is produced by a program which essentially tests every possible simple stemma that could relate these manuscripts' texts and finds the one that minimises the sum of the error values when the frequency of each real text combination is compared with the frequency of that same combination implied by the modelled stemma.

The comparison of the stemmata in Figure 10 and Figure 11 shows that the mathematics do straightforwardly produce a credible, though slightly differing, stemma. Apart from the absence of GA 1689 there are three differences between my best simple stemma and the Lakes' stemma: GA 69 has moved round

to group with 983; GA 543 and GA 826 have changed places; GA 828 has moved to link in at assumed text (e). In order to avoid any confusion with the Lakes' stemma, in the following analysis the identities of the assumed texts have been altered using the lower-case letters j and following. These identities are, of course, simply conventions and we may assign them as we wish. Equally the assumed text (x) is omitted as the analysis cannot create or deal with an assumed text linked to only two other texts, as its content is then indeterminate.

The Lakes do not offer any 'scale' showing how closely related these texts and groups are to each other, but the Probability Structure Analysis result offers such information in the form of stem lengths which give the probable numbers of changes on each stem of Figure 12. This stemma was constructed by a program only from a list of frequencies of the text combinations derived from my counting of variations collated from the Münster transcriptions. Essentially, every possible stemma was considered and the one shown below had the lowest error figure, and therefore the highest probability of explaining the data offered. Thus, the stemma produced depends not at all on any subjective evaluation but wholly on the combination frequencies offered to the program.

In the stemma of Figure 12, the error figure and probability have been reduced to two decimal places. The physical length of each stem roughly mirrors the mean number of changes on that stem, which is given by an attached figure, rounded to a whole number. In later stemmata, where there are partial stems, the width of each dotted line shows very approximately the proportion being represented with a figure appended. Dotted partial stems indicate nothing by their length, as these must be adapted for clarity in presenting the diagrams.

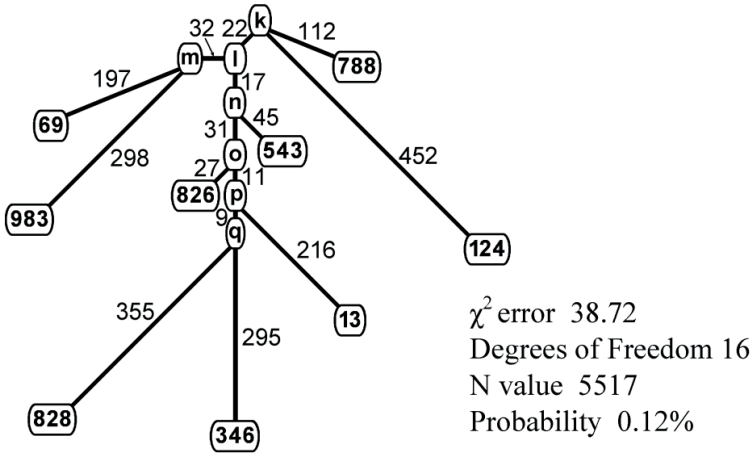

Figure 12

Based on this information, I suggest that while GA 69, 543, 826 and GA 828 have moved in comparison to the Lakes' stemma they have not in fact moved very far. Figure 12 is the optimum simple stemma from these data but has a trivial probability of 0.12%, suggesting it fails to take into account one, or more likely several, conflated texts. As claimed above, Probability Structure Analysis has some ability to account for these conflate texts. In a limited space, I must simply show what I believe at present to be the best analysis that takes conflate texts into account. In many cases a variety of interpretations are possible, but I have omitted these discussions. This is very much work in progress.

How do we find where conflate texts might be? The lack of a conflate text will be indicated by a high error figure for some particular text combination(s). I do not give all 105 combination frequencies which are greater than zero for reasons of space, but Table 6 has the eleven error figures greater than 1.0, ordered by the size of the error. These account for over 90% of the combinations considered.

Combination	Real	Calculated	Error
124, 543, 788, 826	7	0.06986	6.86098
13, 124	7	13.53709	6.09732
69, 124, 543, 788, 826	5	0.49979	4.05037
13, 69, 543, 788, 826, 828	6	1.08104	4.0327
69, 124, 788	6	1.48845	3.39234
124, 346	27	18.83076	2.47172
13, 124, 346, 543, 826, 828	5	1.69082	2.19014
13, 346	14	9.11794	1.70247
69, 788	6	2.98578	1.51426
13, 69, 346, 543, 788, 826, 828	25	19.07667	1.40344
13, 69, 124, 346, 543, 788, 826	11	14.67992	1.23107

Table 6. Error figures derived from Figure 12

By adding partial stems, we can more closely draw together groups which are at present separate but are fragmented. The full expression of the first combination is: 124, 543, 788, 826 :: 13, 69, 346, 828, 983. The left part of the combination contains texts which are contiguous through (k), (l), (n) and (o), uninterrupted by any assumed text connected to a real text not in the group, whereas the right part is fragmented into two groups: Fragment 1 = GA 69, 983 and Fragment 2 = GA 13, 346, 828, which are separated by assumed texts attached to other texts: (l), (n) and (o), as in Figure 12. Of course, we could take the group GA 13, 69, 346, 828, 983 as contiguous through texts (m), (l), (n), (o), (p) and (q) but this would force the remaining texts into three fragments: (k), GA 788 and GA 124; GA 543 on its own and GA 828 on its own. I do not see how this latter fragmentation could be resolved.

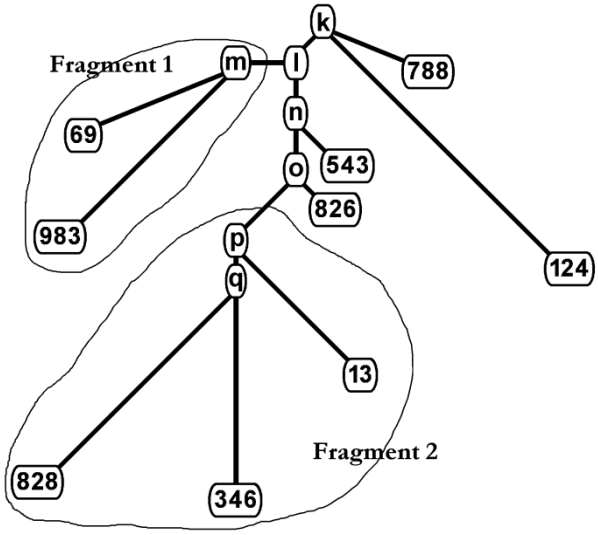

Figure 13

It is the separated groups that we relate together, as in Figure 14. A new stem, (r) to (p), is added in the manner of Figure 8, with partial links (m) to (r), and (o) to (r).

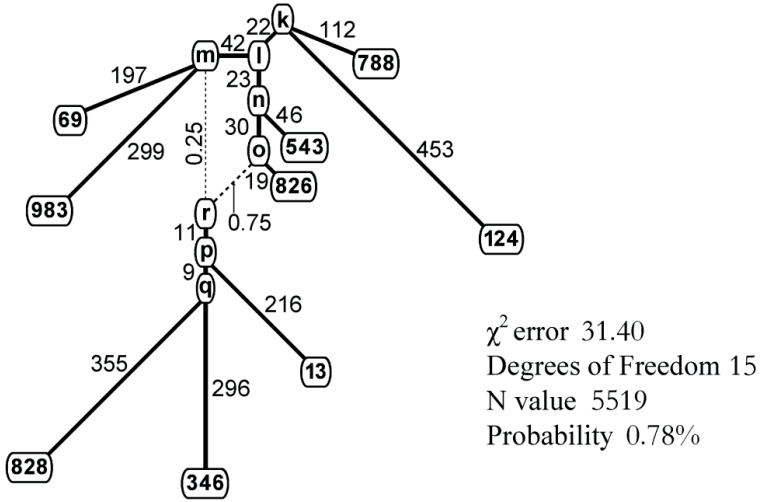

Figure 14

The Chi-squared error figure is reduced by 7.32 to 31.40. One stem, (o) to (p), is lost, a new stem (r) to (p) is added, and a pair of partial stems are added, which count as one new stem since the two parts of the partial stems are fully linked, not independent. The result is 16 + 1 - 1 - 1 = 15 degrees of freedom. The probability is still small at 0.78%. The N value is almost unaltered. The combination 124, 543, 788, 826 :: 13, 69, 346, 828, 983 which has a real frequency of 7 and, from the stemma in Figure 12, had a calculated frequency of 0.06986 with an error of 6.86098, now has a calculated frequency of 7.22752 and an error figure of 0.00740, so is no longer listed among those with error figures in excess of 1.0. Table 7 gives the new list of the four combinations (with error figures greater than 3.0, this time to save space) again listed by size of error.

Combination	Real	Calculated	Error
13, 124	7	13.54445	6.11854
13, 69, 543, 788, 826, 828	6	1.08334	4.02893
69, 124, 543, 788, 826	5	0.75973	3.59597
69, 124, 788	6	1.49154	3.38771

Table 7. Error figures derived from Figure 14

The two combinations with the highest error figures here cannot be used to modify the stemma. The first, GA 13, 124 :: 69, 346, 543, 788, 826, 828, 983, has a higher calculated frequency than the real frequency. The stemma can only be modified now by adding a stem, full or partial, which necessarily increases the appropriate frequency, but cannot decrease it. The second combination, GA 13, 69, 543, 788, 826, 828 :: 124, 346, 983, seems to be resolvable as the calculated frequency is less than the real frequency, but it requires a modification to increase the frequency of agreement between three texts which are fragmented in the stemma. This would require three extra stems, linking GA 124 to GA 346, linking GA 124 to GA 983 and also linking GA 346 to GA 983. This real data from nearly the whole of Mark, where the variations have only two forms, is the first real data that has produced such situations and is on the long list of things requiring further work. Hence, we work with the third

combination which is similar to that we used to modify Figure 12: GA 69, 124, 543, 788, 826 :: 13, 346, 828, 983. This time the unlinked fragments are GA 13, 346, 828 and GA 983, whereas in Figure 12 and Figure 13 one fragment contained both GA 69 and GA 983. There is already a link from (r) to (m); so, rather than add another link, the link from (r) to (m) is moved to a new assumed text (s) somewhere between (m) and GA 983. The optimized result is shown in Figure 15.

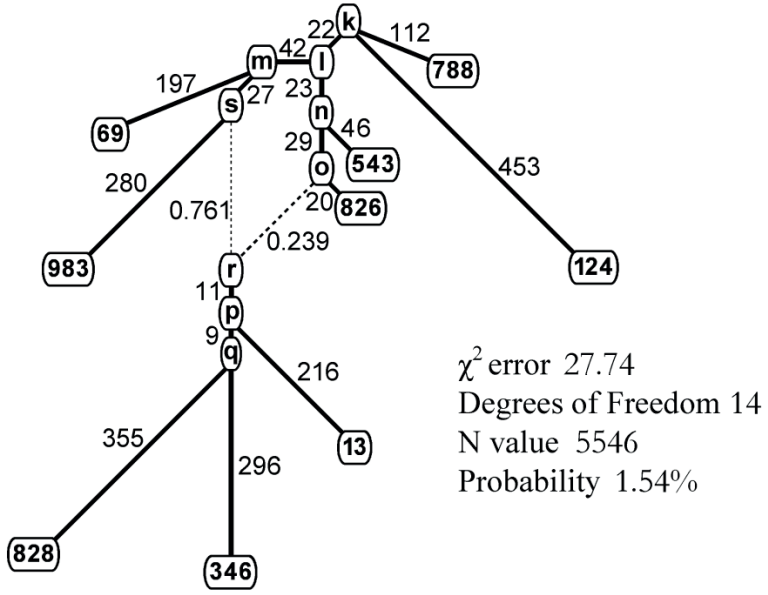

Figure 15

Again, the error figure is less, while the degrees of freedom are reduced by one as a new stem from (m) to (s) is added, the N figure increases slightly and the probability is now 1.54%. There are slight changes in the lengths of a few stems; in fact every length has changed slightly, hidden by the rounding of the figures to whole numbers. The combination GA 69, 124, 543, 788, 826 :: 13, 346, 828, 983 which has a real frequency of 5 and had, from the stemma in Figure 14, a calculated frequency of 0.75973 and an error of 3.59597, now has, from the stemma in Figure 15, a frequency of 4.00212 and an error of 0.19915. Table 8 lists the

error figures resulting from the stemma shown in Figure 15 which are greater than 2.0.

Combination	Real	Calculated	Error
13, 124	7	13.49845	6.03284
13, 69, 543, 788, 826, 828	6	1.07187	4.04774
69, 124, 788	6	1.51347	3.35483
124, 346	27	18.76370	2.51247
13, 124, 346, 543, 826, 828	5	1.68387	2.19934

Table 8. Error figures derived from Figure 15

The combinations with the two highest errors still do not resolve so we attend to the next: GA 69, 124, 788 :: 13, 346, 543, 826, 828, 983. The first part of the combination fragments into GA 69 against GA 124 with GA 788, so partial links are put in from (k), linked to GA 124 and GA 788, and from (s) to a new assumed text (t) which feeds GA 69 directly, as in Figure 16. By adding a partial stem (t) to (s) but eliminating a full stem, from (l) to (m) the degrees of freedom remain at 14. Probability is now at 4.17%.

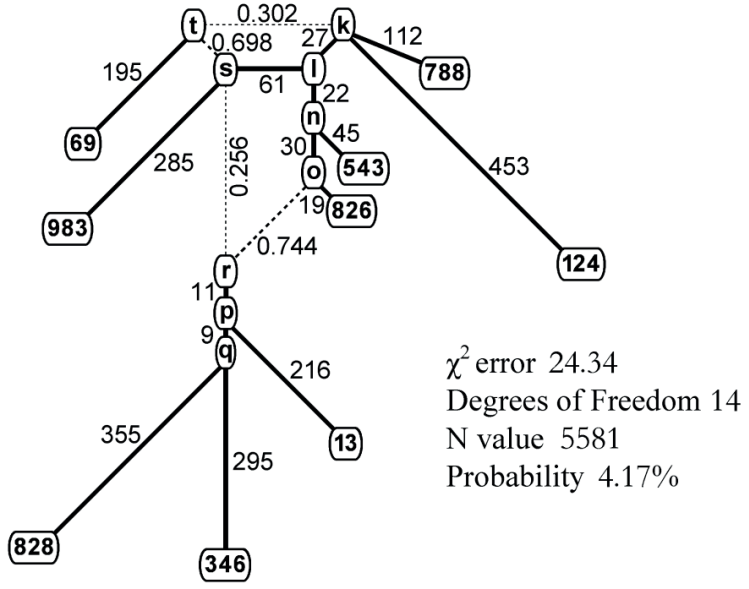

Figure 16

The combination GA 69, 124, 788 :: 13, 346, 543, 826, 828, 983 with a real frequency of 6 now has a calculated frequency of 6.95246 and the error is reduced from 3.35483 to 0.15120. Table 9 lists the error figures resulting from the stemma shown in Figure 16 which are greater than 2.0.

Combination	Real	Calculated	Error
13, 124	7	13.44853	5.9405
13, 69, 543, 788, 826, 828	6	1.05678	4.07257
124, 346	27	18.67375	2.56765
13, 124, 346, 543, 826, 828	5	1.62425	2.27914
13, 124, 346, 543, 826, 828	5	1.68387	2.19934

Table 9. Error figures derived from Figure 16

The first and second of these combinations remain unworkable but the third, GA 124, 346 :: 13, 69, 543, 788, 826, 828, 983 is resolved by linking the fragments GA 124 and GA 346. This is achieved by making complimentary partial links from (q) and from GA124 to a new assumed text (u), which reduces the degrees of freedom to 13, as in Figure 17.

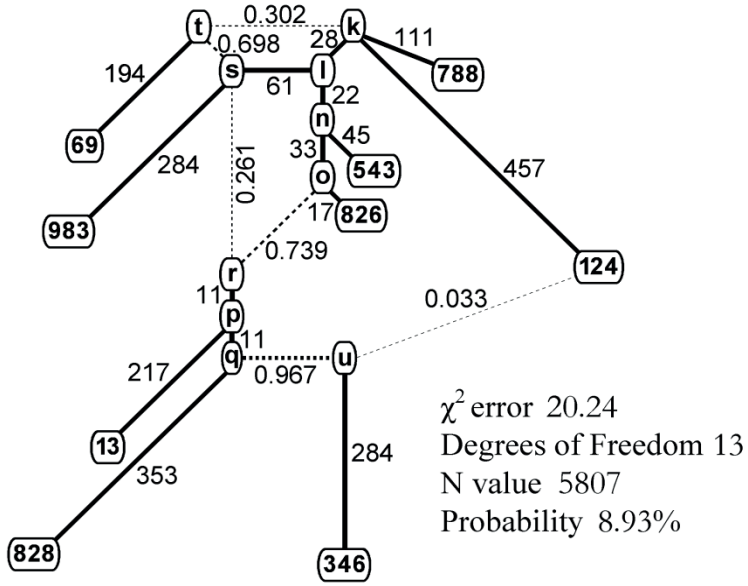

Figure 17

The error figure is now reduced to 20.24, the N value is 5807 and the probability 8.93%. The combination GA 124, 346 :: 13, 69, 543, 788, 826, 828, 983 has a real frequency of 27. In the stemma in Figure 16 the combination had a calculated frequency of 18.67375 and an error of 2.56765, but in the stemma in Figure 17 it has a calculated frequency of 28.26250 and an error figure of 0.05903. Table 10 lists the error figures resulting from the stemma in Figure 17 which are greater than 1.0, covering 86% of the total error.

Combination	Real	Calculated	Error
13, 124	7	12.79122	4.79117
13, 69, 543, 788, 826, 828	6	1.52576	3.33647
13, 124, 346, 543, 826, 828	5	1.58876	2.32732
13, 346	14	8.46735	2.18644
13, 69, 346, 543, 788, 826, 828	25	17.81327	2.06596
13, 69, 124, 543, 788, 826, 828	16	11.11718	1.49012
69, 788	6	3.29337	1.22098

Table 10. Error figures derived from Figure 17

The first two combinations remain irresolvable so we work with the next: GA 13, 124, 346, 543, 826, 828 :: 69, 788, 983. It is the second grouping that is fragmented. GA 983 and GA 69 are closely linked but GA 788 does not link to them closely. The solution is to make GA 788 conflate from (s) and its existing root (k) with a new assumed text is added (v) linking directly to GA 788 as in Figure 18.

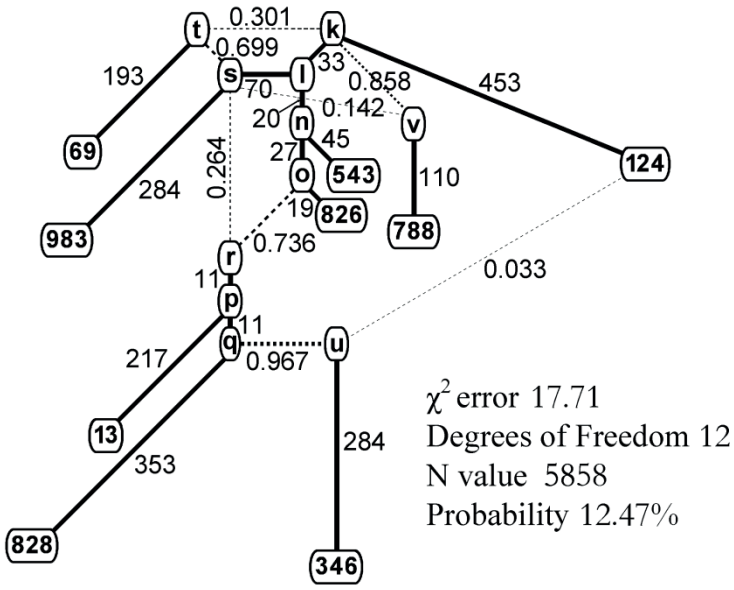

Figure 18

Interestingly, GA 69 now has the same partial links, to (k) and to (s), as GA 788, though in different proportions. The extra partials reduce the degrees of freedom by one to 12. The N value is 5858, and the error figure 17.71 so that the probability is now 12.47%. The combination GA 13, 124, 346, 543, 826, 828 :: 69, 788, 983, which has a real frequency of 5, had a calculated frequency of 1.58876 and an error figure of 2.32732 from the stemma in Figure 17, but from the stemma in Figure 18 it has a calculated frequency of 5.08986 and an error figure of 0.00162. Table 11 lists the error figures from the stemma in Figure 18 which are greater than 1.0.

Combination	Real	Calculated	Error
13, 124	7	12.6798	4.60859
13, 69, 543, 788, 826, 828	6	1.50038	3.37442
13, 346	14	8.40278	2.23777
13, 69, 346, 543, 788, 826, 828	25	17.63076	2.17223
13, 69, 124, 543, 788, 826, 828	16	11.01962	1.55026
69, 788	6	3.41611	1.11274

Table 11. Error figures derived from Figure 18

Again, the first two combinations are irresolvable so we look at GA 13, 346 :: 69, 124, 543, 788, 826, 828, 983. GA 346 is already a conflate from (q) and GA 124, so the simplest approach is to add a further link from GA 13 to (u) as in Figure 19. I attempted to resolve this fragmentation by juggling the full stems around GA 13, GA 828 and GA 346 but found no simple solution.

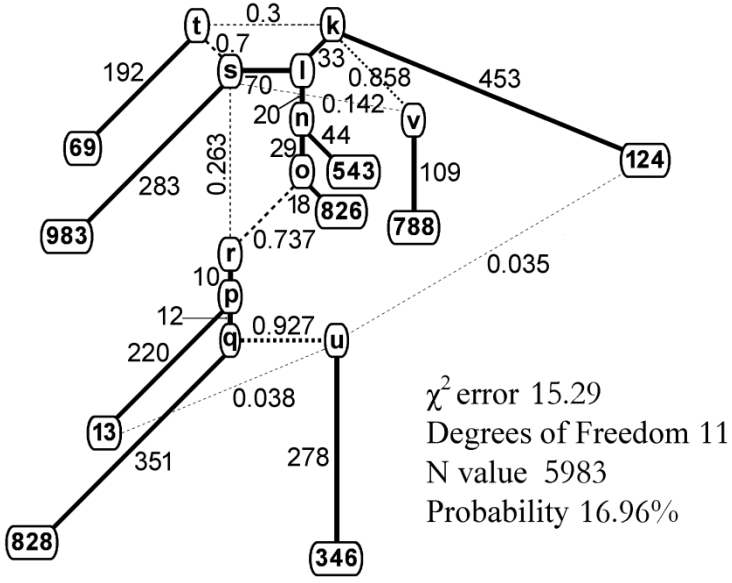

Figure 19

The extra partial link reduces the degrees of freedom by one to 11. The combination 13, 346 :: 69, 124, 543, 788, 826, 828, 983 had a real frequency of 14 and a calculated frequency of 8.40278, giving an error of 2.23777 from the stemma in Figure 18, but it has a calculated frequency of 13.99175 giving an error of 0.00000 in Figure 19. Table 12 lists the error figures from the stemma in Figure 19 which are greater than 1.0.

Combination	Real	Calculated	Error
13, 124	7	12.22954	3.90687
13, 69, 543, 788, 826, 828	6	1.46997	3.42019
13, 69, 346, 543, 788, 826, 828	25	17.2733	2.38808
13, 69, 124, 543, 788, 826, 828	16	10.60811	1.81703
69, 788	6	3.3355	1.18326

Table 12. Error figures derived from Figure 19

Again, the first two combinations are irresolvable so we look at GA 13, 69, 346, 543, 788, 826, 828 :: 124, 983. The pair GA 124 and GA 983 must be fragmented to fit the stemma in Figure 19. Figure 20 shows GA 983 made conflate and derived from a new assumed text (w) fed by partial links from (s) and GA 124.

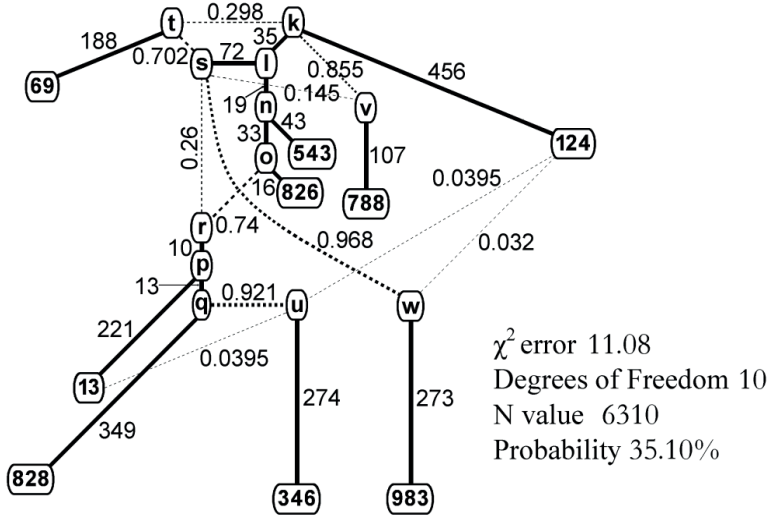

Figure 20

The stemma in Figure 19 gave a calculated frequency of 17.27330 and an error of 2.38808 to the combination GA 13, 69, 346, 543, 788, 826, 828 :: 124, 983, which had a real frequency of 25, but the stemma in Figure 20 gives a frequency of 26.40976 and an error of 0.07950. Table 13 lists the error figures resulting from the stemma in Figure 20 which are greater than 1.0.

Combination	Real	Calculated	Error
13, 124	7	11.44507	3.90687
13, 69, 543, 788, 826, 828	6	2.28685	2.29791
13, 69, 124, 543, 788, 826, 828	16	9.70306	2.47822
69, 788	6	3.30054	1.21451

Table 13. Error figures derived from Figure 20

Again, the first two combinations are irresolvable so we look at GA 13, 69, 124, 543, 788, 826, 828 :: 346, 983. Figure 21 shows a new partial link between GA 983 and (u) which feeds GA 346.

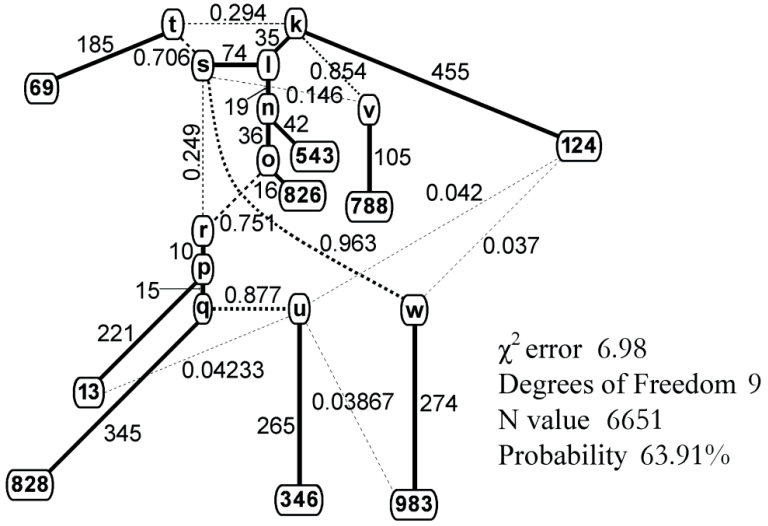

Figure 21

The new link decreases the degrees of freedom by one. The combination 13, 69, 124, 543, 788, 826, 828 :: 346, 983 has a real frequency of 16. The stemma in Figure 20 gave a frequency of 9.70306 and hence an error of 2.47822. The stemma in Figure 21 gives a frequency of 17.10938 and an error of 0.07692. Table 14 lists the error figures resulting from the stemma in Figure 21 which are greater than 1.0.

Combination	Real	Calculated	Error
13, 124	7	10.85372	2.12160
69, 788	6	3.15786	1.34629
13, 69, 543, 788, 826, 828	6	3.42002	1.10938

Table 14. Error figures derived from Figure 21

Again, the first and third combinations are irresolvable so we look at GA 69, 788 :: 13, 124, 346, 543, 826, 828, 983. A link is put in between GA 788 and (t) giving the arrangement in Figure 22. There is the alternative of linking these manuscripts by putting the link between GA 69 and (v) but this yields a slightly lower probability.

After the first run of this stemma, I made two changes. First, I reduced the stem between (n) and (o) to zero, as the first run of the stemma gave it a length of two changes. I rejected this length because I doubt a text could be copied with only two changes and also because removing the stem—that is reducing this stem to zero length—increases the probability figure slightly since the degrees of freedom are increased thereby. Second, I tried putting the two small links to GA 124, from (u) and from (w), to a point on the stem GA 124 to (k) at an assumed text I called (j). This gave a better probability and fitted better with the Lakes' understanding of the history of the text in Figure 11, which I accepted in my rendering of the optimum simple stemma in Figure 10. My reason for doing this is dealt with later in the chapter. The final form of the stemma of Mark in Family 13, at this stage of my ongoing studies, is in Figure 22. I have twisted the stemma somewhat in the hope of making it more legible.

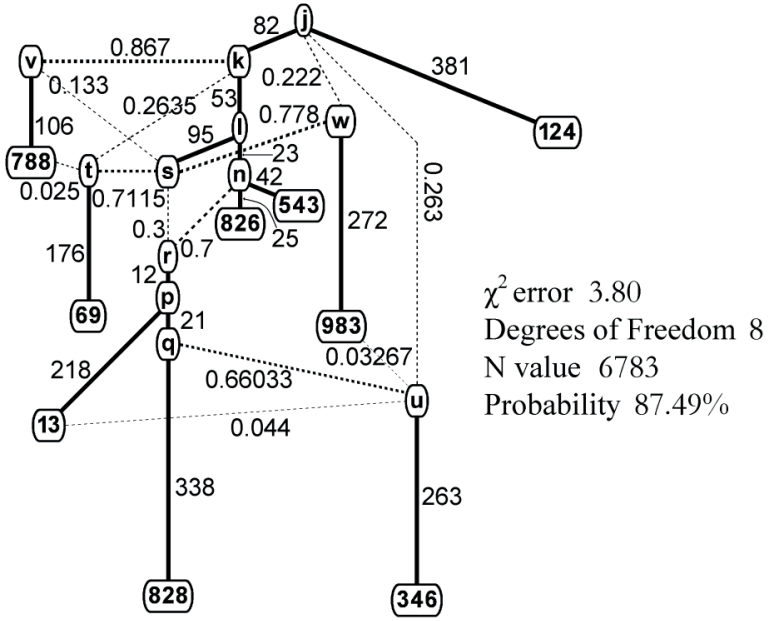

Figure 22

The combination GA 69, 788 :: 13, 124, 346, 543, 826, 828, 983, which has a real frequency of 6, had, according to the stemma in Figure 21, a calculated frequency of 3.15786 and an error of 1.34629, but according to the stemma in Figure 22 has a calculated frequency of 5.91779 giving an error figure of 0.00113.

Several general remarks are proper here. As partial stems have been added, the probability that the stemma is a credible explanation of the relationships between the manuscript texts has increased (it has, of course, to be the right stem in the right place) although degrees of freedom have decreased making the achievement of a high probability more difficult. The N figure has steadily increased, but a discussion of why this should be is well beyond the scope of this article. The probability for the stemma in Figure 22 is quite high at 87.49%. Perhaps this is too high. Statisticians can be suspicious of correlations that are too good: anything above 97.5% is beyond credibility. The difficulty here is that one could go on adding partial links to the stemma until workable ones ran out; but this might simply be trying to resolve

chance agreements into partial links. I cannot see, yet at least, any difference between combinations that need fragmenting in order to fit our working stemma because some part of the texts' history has not yet been taken into account and those which naturally arise from chance agreements and chance reversals of the text. But I do need to find some clear indication of where to stop; this is work in progress.

The stemma has become complicated, but is it still, in any sense, basically the same stemma as that in Figure 10? In Figure 23 I have taken the stemma in Figure 22 and removed all the smaller partial links and made the strongest partial link in each case a full link. Allowing for distortions of length and twisting of stems the reader can see that Figure 10 and Figure 23 are almost the same stemma, with the one exception that the stems from GA 543 and GA 826 now join the rest of the stemma at the same point.

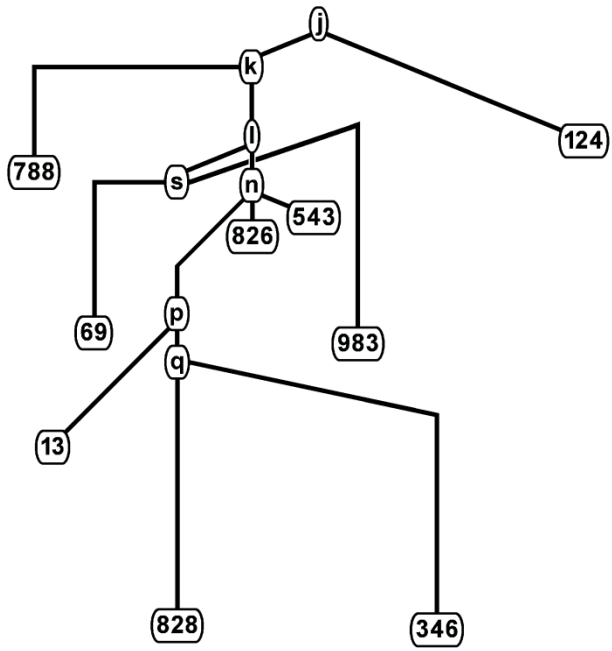

Figure 23

As I progressed in adding partial stems, two text combinations could not be construed and so were ignored in the calculations. That is: GA 13, 124 :: 69, 346, 543, 788, 826, 828, 983 which had an error of 6.09732 in the optimum simple stemma and GA 13, 69, 543, 788, 826, 828 :: 124, 346, 983 which had an error of 4.03270 in the optimum simple stemma. The appropriate error figures have generally reduced as the complication of the stemma has gone forward and, in the stemma in Figure 22, stand at 1.84130 (the highest remaining error) and 0.01890, respectively. As the computer tries at each stage to modify the stem lengths and proportions of partial stems every combination frequency is considered so that, inevitably, each error figure will be altered. The stemma is in some measure an interconnected whole. In preparing Figure 22, I pointed out that I had rerun the stemma with the stem (n) to (o) removed because it was unreasonably small. What then should we make of the very small partial links that we see in the later stemmata of Family 13?[9] Would the result be better by ignoring and removing them? The answer is no, since by doing so the error figure increases considerably. But can we sensibly imagine a situation where a copyist adds just a few verses from a different source to that from which nearly all the text is copied?

As mentioned above, a scribe can need to deal with copies which have missing portions of text: indeed, the manuscripts that I have considered for the stemma of Family 13 are defective: GA 13 is missing 1:21 to 1:45a; GA 543 is missing 8:4b to 8:28a; 826 is missing 12:3b to 12:19a. Since Mark has 678 verses these missing sections amount to, as decimal proportions, 0.036, 0.035 and 0.024, respectively. If someone used one of these manuscripts as their exemplar, they would need to supplement the text with these small proportions from elsewhere. It is therefore conceivable, but not of course proved, that this has happened in copying texts in the stemmata in this article. The exact proportion would depend on the number of verses to a folio in the copy text and the number of folios missing. Equally, in comparing any two manuscript texts, differ-

[9] In Figure 22: links 788 to (t) = 0.025; 13 to (u) = 0.044; 983 to (u) = 0.03267. Further small partial links occur in Figures 24 to 26.

ences are not evenly distributed, hence our observation here is a very general one. Where proportions are more substantial it seems simple enough to account for them by the scribe exchanging copy text part way through or a scribe fairly industriously, but only partially, 'correcting' the copy text. Where a number of partial stems converge, as with assumed text (u) in Figure 22, it must be remembered that some, perhaps many, manuscripts are missing from our knowledge and the complex situation shown may be all we have left of an extensive history of copying.

I am struck by the difference of the stem lengths in Figure 22; several in the twenties and others in the three hundreds. Perhaps the longer stems are testimony to exceptionally deficient work, but I suspect it is much more likely that these stems are records of many competent individual copying events. We are saddened by what has been lost, but I trust we may recover some of this by effective implementation of good theoretical work

The further issue that needs tidying up is the question of the point of origin of my stemmata. My method of determining which text in the stemma has the earliest text has been to include NA28 in the collation. NA28 has, as I understand it, the current best critical text of Mark since the ECM was not yet available at the time of writing. Of course, NA28 is not a manuscript text but a scholarly text. However, the justification for its inclusion in this part of the study is simply that it works. I have not included NA28 in the main analysis of this article as including it with the whole family introduces complications that I have not yet mastered and the discussion of which extends far beyond the principal aim of this study; and crucially, of course, NA28 is not a member of Family 13. However, for my purposes here, I introduce a worked stemmata of NA28, GA 69, GA 124, GA 788 and GA 983 which are those manuscripts nearest to the point where the Lakes place their earliest text as in Figure 1 and Figure 11. Only five texts are fairly easy to deal with, though I am not suggesting that the results here are necessarily significant in any wider context. Figure 24 shows the optimum simple stemma as determined by the analysis program simply on the basis of the manuscript combinations. However, since the stemma has a poor probability, I have pursued a better stemma to provide a more secure result.

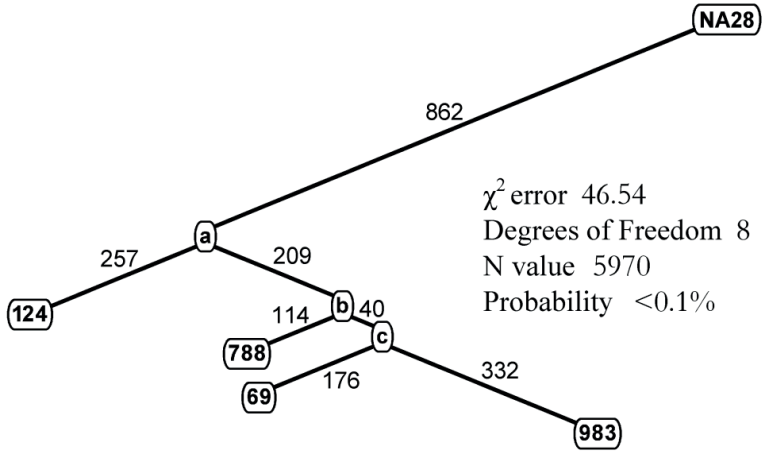

Figure 24

Table 15 lists all the combinations, real frequencies, calculated frequencies, and corresponding error figures from the stemma in Figure 24.

Combination	Real	Calculated	Error
NA28 :: 69, 124, 788, 983	730	717.85055	0.20220
NA28, 69 :: 124, 788, 983	14	22.32308	4.94812
NA28, 69, 124 :: 788, 983	10	10.53649	0.02878
NA28, 69, 788 :: 124, 983	20	13.13079	2.35930
NA28, 69, 124, 788 :: 983	250	248.84887	0.00530
NA28, 124 :: 69, 788, 983	181	185.59209	0.11650
NA28, 124, 788 :: 69, 983	35	39.78777	0.65494
NA28, 788 :: 69, 124, 983	35	15.84529	10.48294
69 :: NA28, 124, 788, 983	145	129.79340	1.59476
69, 124 :: NA28, 788, 983	6	7.44259	0.34685
69, 124, 788 :: NA28, 983	34	42.50416	2.12708
69, 788 :: NA28, 124, 983	28	13.44141	7.56973
124 :: NA28, 69, 788, 983	209	215.25725	0.18734
124, 788 :: NA28, 69, 983	32	10.35936	14.63491
788 :: NA28, 69, 124, 983	74	83.75375	1.28562

Table 15. Error figures from Figure 24

The combination GA 124, GA 788 :: NA28, GA 69, GA 983 has the highest error figure. GA 124 and GA 788 are already closely linked, so the fragment NA28 and the fragment (c) with GA 69 and GA 983 are connected by partial links to a new assumed text (d) as in Figure 25.

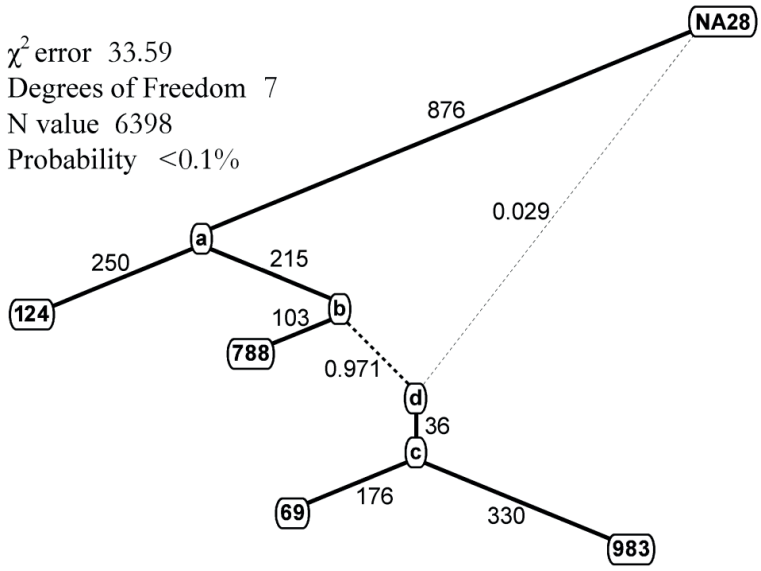

Figure 25

The error figure is reduced but the probability is still tiny.[10] The highest error from the stemma in Figure 25 is for NA28, GA 788 :: GA 69, GA 124, GA 983 with a real frequency of 35, a calculated frequency of 13.27473 and an error of 13.48535. This error is reduced by separating GA 788, with an assumed text (e) fed by two partial stems linked to NA28 and (b) as in Figure 26.

[10] In this section, to save space, I will not keep offering lists of error figures as this is not part of the main presentation.

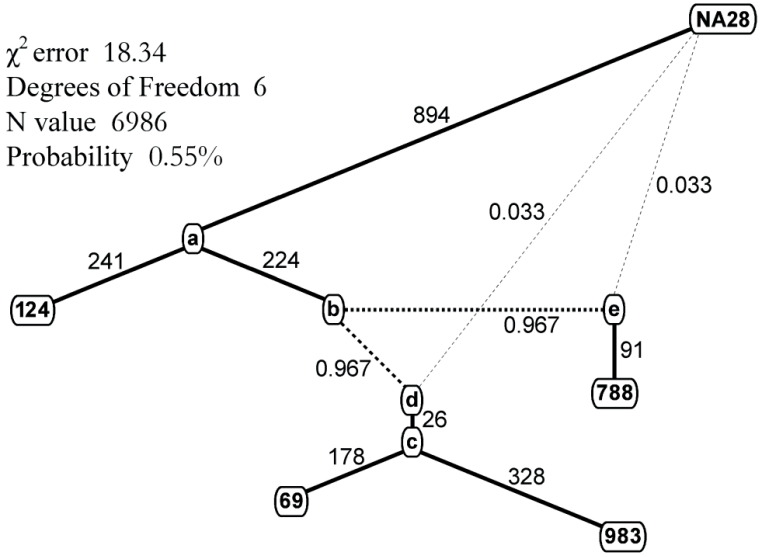

Figure 26

The error is again reduced and the probability just noticeable at 0.55%. The highest error now from the stemma in Figure 26 is for combination GA 69, GA 788 :: NA28, GA 124, GA 983 with a real frequency of 28, a calculated frequency of 11.18141 and an error of 10.10232. This is dealt with by adding a third assumed text (f) linking to 69 with partial links to (c) and GA 788, as in Figure 27.

The probability is now high and we need proceed no further. So, what do we conclude as to the point of origin of the completed stemma? Consistently through the four stemmata in this last section, NA28, representing for us at least a very early form of Mark, joins the family stemma between GA 124 and (b), the point at which strong links to GA 788 and to the common ancestor of GA 69 and GA 983 come together. In these stemmata the proportions of the two sections between GA 124 and (b) are roughly equal. In the final stemma for our present analysis in Figure 22, these lengths between GA 124 and (j), and (j) and (k), are much less equal. However, looking at Figure 21, the origin of the two very small partial links ending at (u) and (w) comes at GA 124, in a stemma with a probability of over 60%. If this coincident point suggests the position of the earliest text, it seems

that wherever it is placed between GA 124 and (k) will offer a fairly high probability. For an accurate positioning, I would suggest that a full analysis with NA28 included with the whole Family is the only certain way forward. But, for the time being, a point between GA 124 and (k) is credible as in Figure 10 and Figure 11.

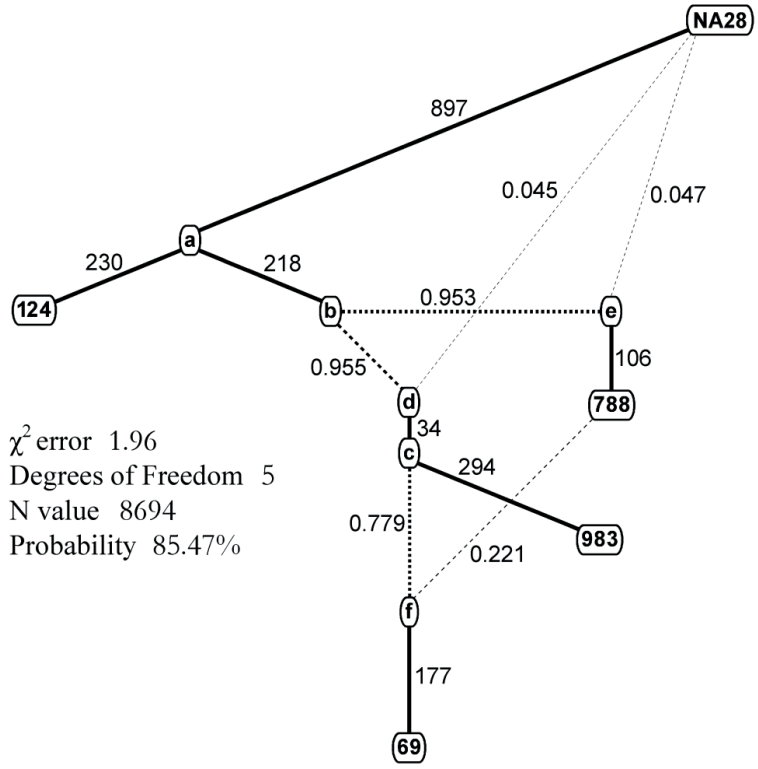

Figure 27

CONCLUSIONS

I hope this study has demonstrated that the underlying conjectures of Probability Structure Analysis, and the as yet imperfect method of applying it, show promise in explaining the relationships between the texts of the manuscripts examined, and that the whole is worthy of further work and consideration. Further research, giving a fairly full examination of the methods and poten-

tial of Probability Structure Analysis, will look more carefully at the relationship between Family 13 and the earliest text. Indeed, I hope to collate in some version of the Byzantine text to see where the link to those forms of the text can be clarified. Equally, Probability Structure Analysis is not restricted to the analysis of closely related groups of manuscript texts but extends to any group of texts, or any groups of texts now only considered distant relatives, provided they are of a single work.