



Digitilizing New Testament textual criticism

Pasi Hyytiäinen
Postdoc researcher
Faculty of Theology



In this presentation

- New Testament textual criticism
 - Nature and purpose of the field
 - Challenges
- Digital techniques used in the field
 - Possibilities and limitations
- My solution to the key challenges of New Testament textual criticism
 - Relate



” The Bible did not fall magically from the clouds. Man created it as a historical record of tumultuous times, and it has evolved through countless translations, additions, and revisions. History has never had a definitive version of the book.”

Sir Leigh Teabing, Da Vinci Code

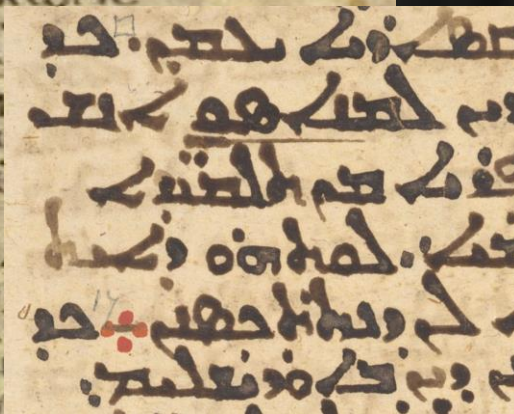
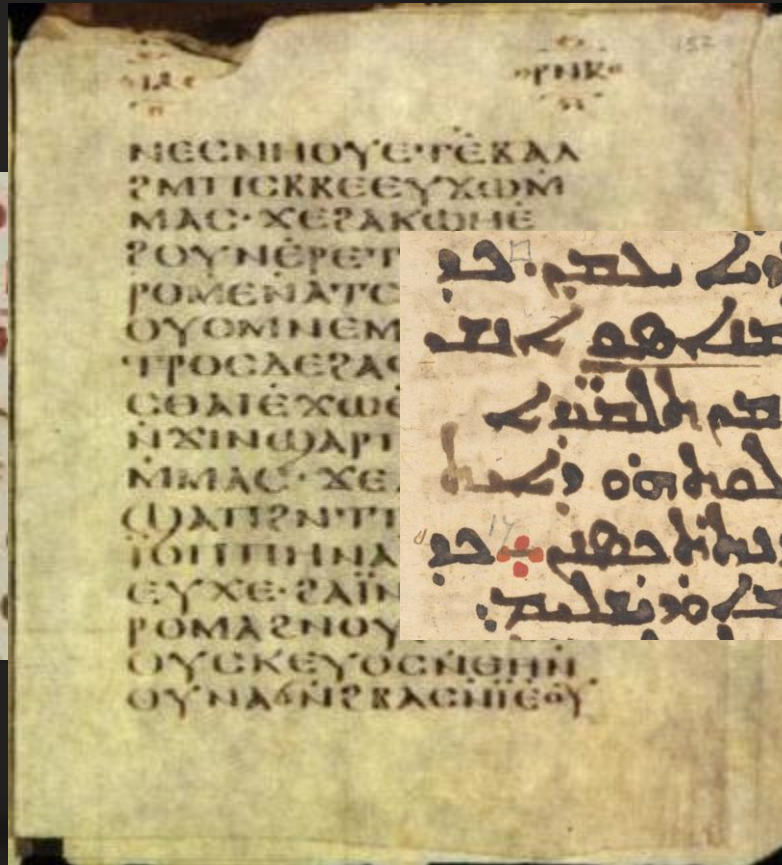
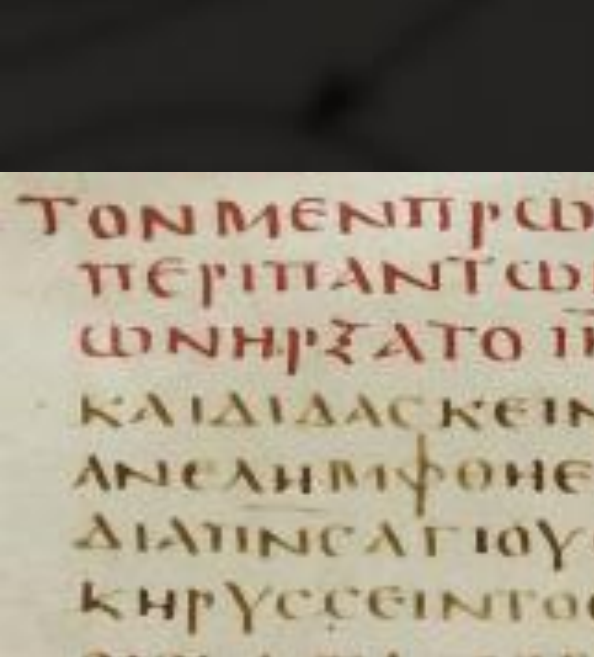


The transmission of the New Testament through centuries

- New Testament was transmitted through the medium of manuscripts for over 1400 years
 - Manuscript = manus scriptus (“written by hand”)
- 27 different manuscript traditions
 - Each text circulated individually during the first Christian centuries
 - Every one of them has a unique history
 - Later assembled in collections of books
 - Gospels, Apostolos
 - Numerous translations
 - Greek, Latin, Coptic, Syriac...



Manuscript traditions of the New Testament





New Testament textual criticism as a discipline

- Textual criticism aims to organize the manuscripts in a meaningful manner
- Traditional task is to reconstruct the oldest possible textual form
 - Based on the manuscript data
 - Comparison of different manuscripts to decide, which represent the oldest text
 - Leads to a critical edition
 - Nestle-Aland (NA)
 - United Bible Societies (UBS)
 - Editio Critica Maior (ECM)




First Challenge

- The sizes of the manuscript traditions
 - Approximately 5700 Greek
 - 10 000 Latin
 - 1500 Armenian
 - 1000 Coptic New Testament manuscripts
- New manuscript discoveries are made each year

- Welcome
- Manuscript Workspace
- ECM
- NT Conjectures
- Forum
- Blog
- Liste
- Transcribing
- Indexing
- Status
- NA28
- Collation
- Help
- About

Full Search

Manuscript Num. **Name** ▾ Clear All 

Current Country: ▾

Place: ▾

Institute: ▾


Shelf Num.: ▾


Indexed Biblical Content ▾ ▾ ▾


Content Overview ▾


Language ▾


Has Feature ▾


Dated To 0 - 1799AD 

Line Count 0 - ∞ 

Columns 0 - ∞ 

Page Height (mm) 0 - 500 

Page Width (mm) 0 - 450 

Folio Count 0 - ∞ 

Has: Images Transcriptions

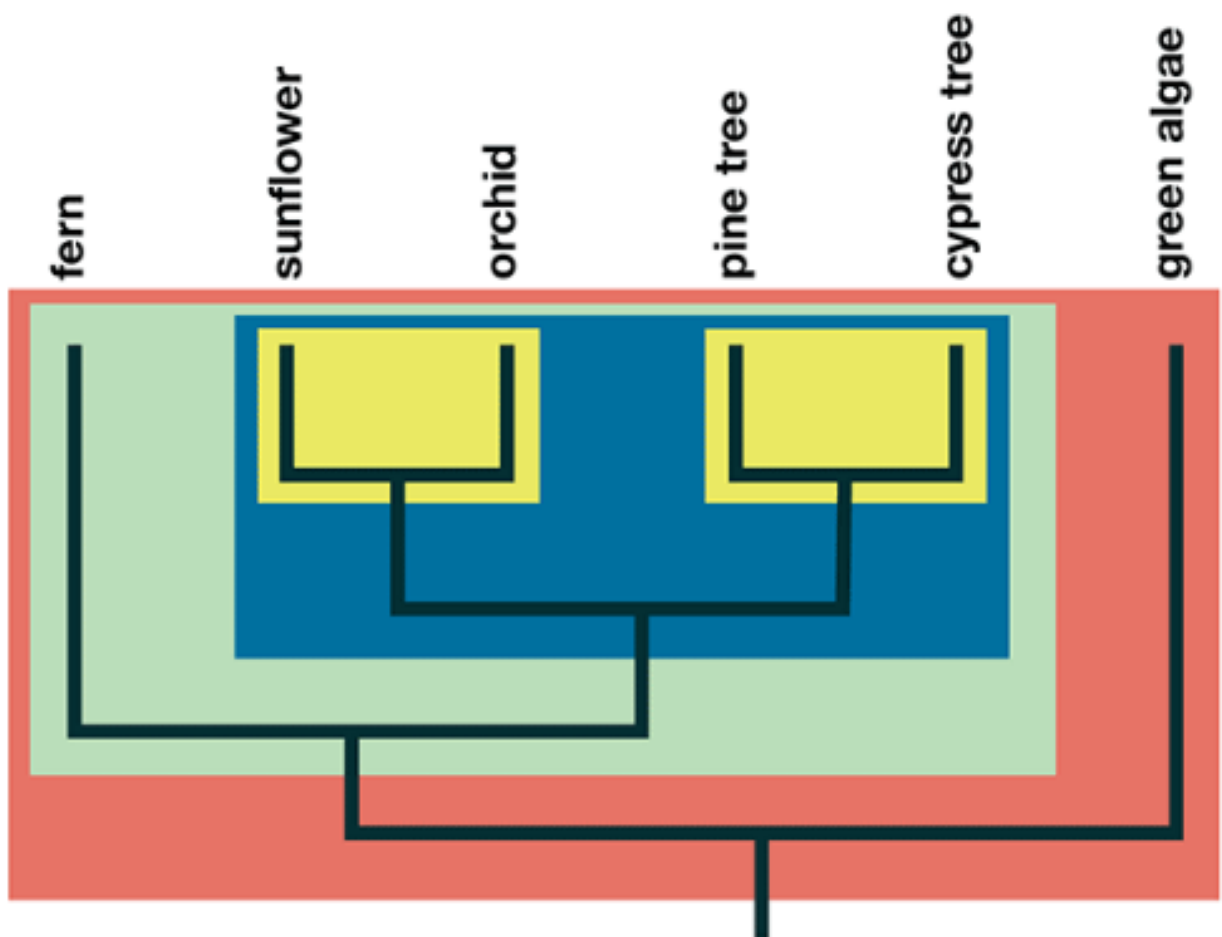
Manuscript Details

Bibliography



Tackling the challenge: Hierarchical classification

- Hierarchical classification has been used to organize the data
 - Johan Bengel 1734 (Carl Linnaeus 1735)
 - Bengel classified manuscripts and Linnaeus natural world in a similar manner

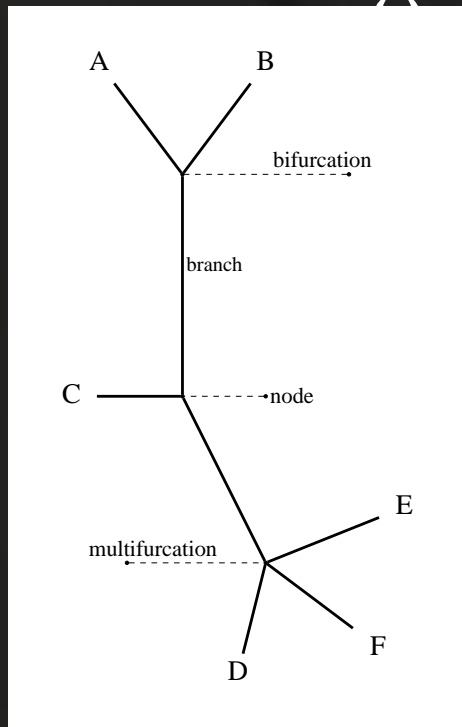


the data

in a similar



Tackling the challenge: Phylogenetic methods



Originally developed for evolutionary biology

parallel phenomena between
qual criticism

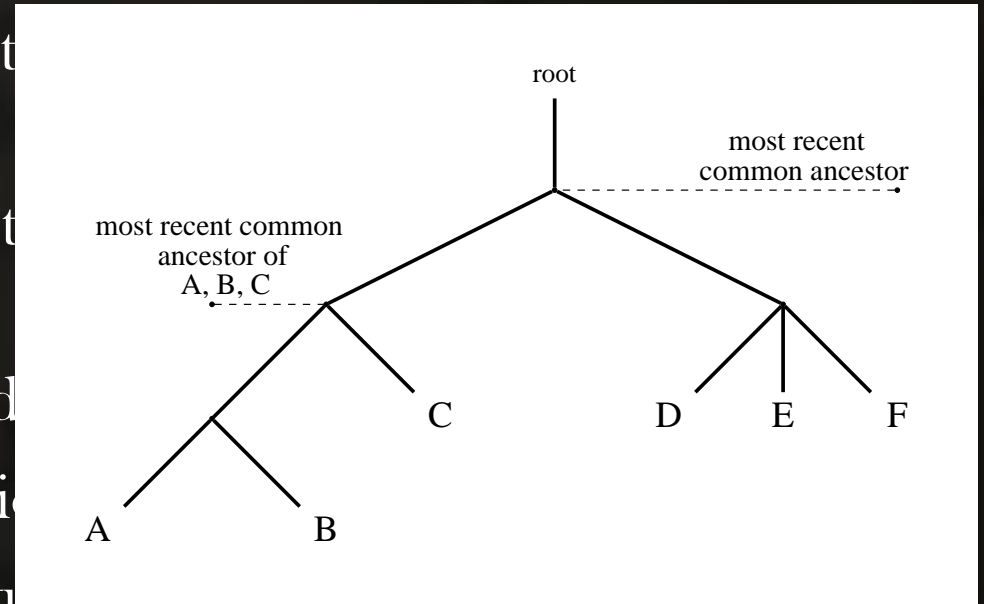
enables the usage of the
udies



es or networks are d

the level of automatic

aster analysis of textual traditions



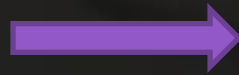
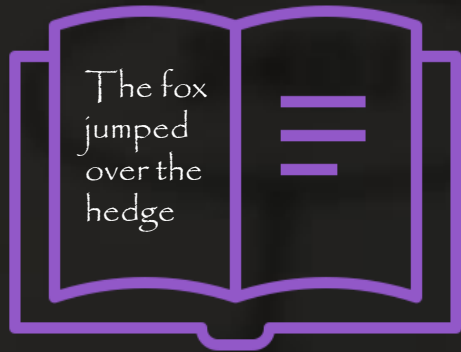


First Challenge

- Considering all manuscript data is impossible when conventional phylogenetic techniques are used
 - Critics are forced to rely on samples
 - Preprocessing work takes too much time
 - Transcriptions
 - Collations
 - Establishing places of variation
 - Encoding variation places



Transcribing process



the fox jumped over the hedge



Collating manuscripts

- A the fox jumped over the hedge
- B -
- C the cat jumped over the fence
- D a man saw that the fox jumped over the hedge
- E a man saw that the fox jumped over the fence



Establishing variation places

Solution 1			Solution 2		
A	–	the fox jumped over the hedge	–	the fox jumped	over the hedge
B	–	–	–	–	–
C	–	the cat jumped over the fence	–	the cat jumped	over the fence
D	a man saw that	the fox jumped over the hedge	a man saw that	the fox jumped	over the hedge
E	a man saw that	the fox jumped over the fence	a man saw that	the fox jumped	over the fence



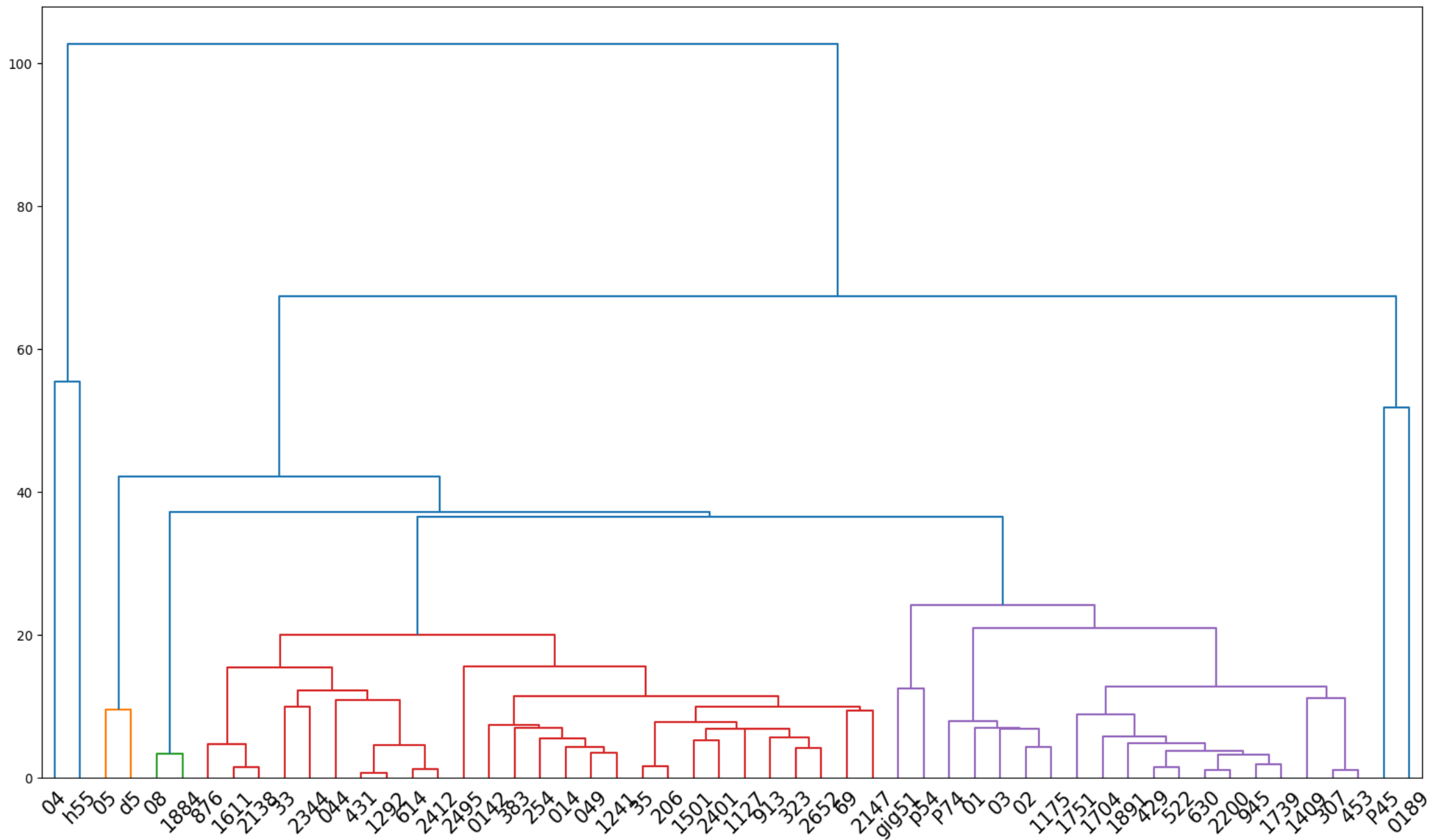
Encoding variation places

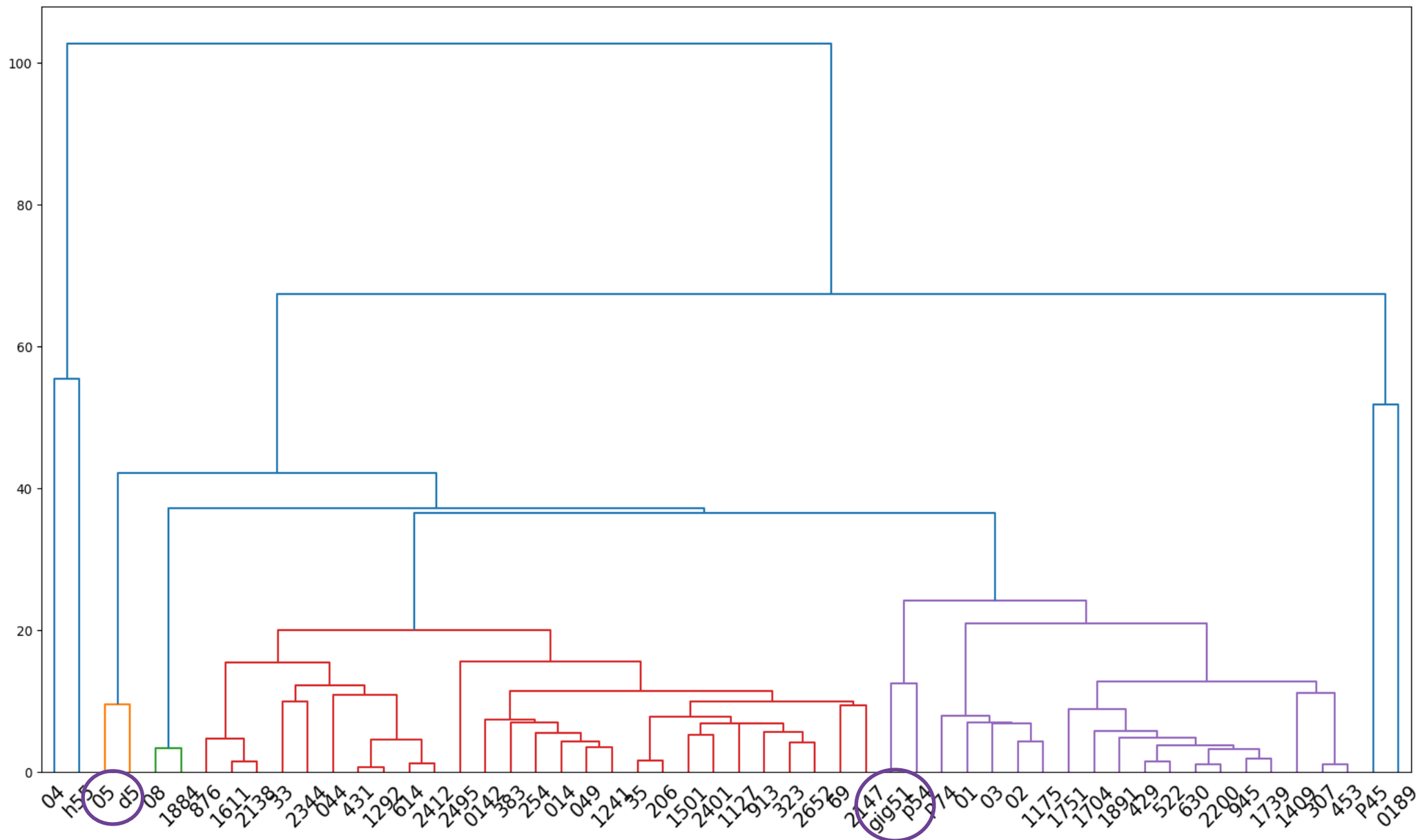
	Solution 1		Solution 2		
A	0	1	0	0	0
B	?	?	?	?	?
C	0	0	0	1	1
D	1	1	1	0	0
E	1	2	1	0	1

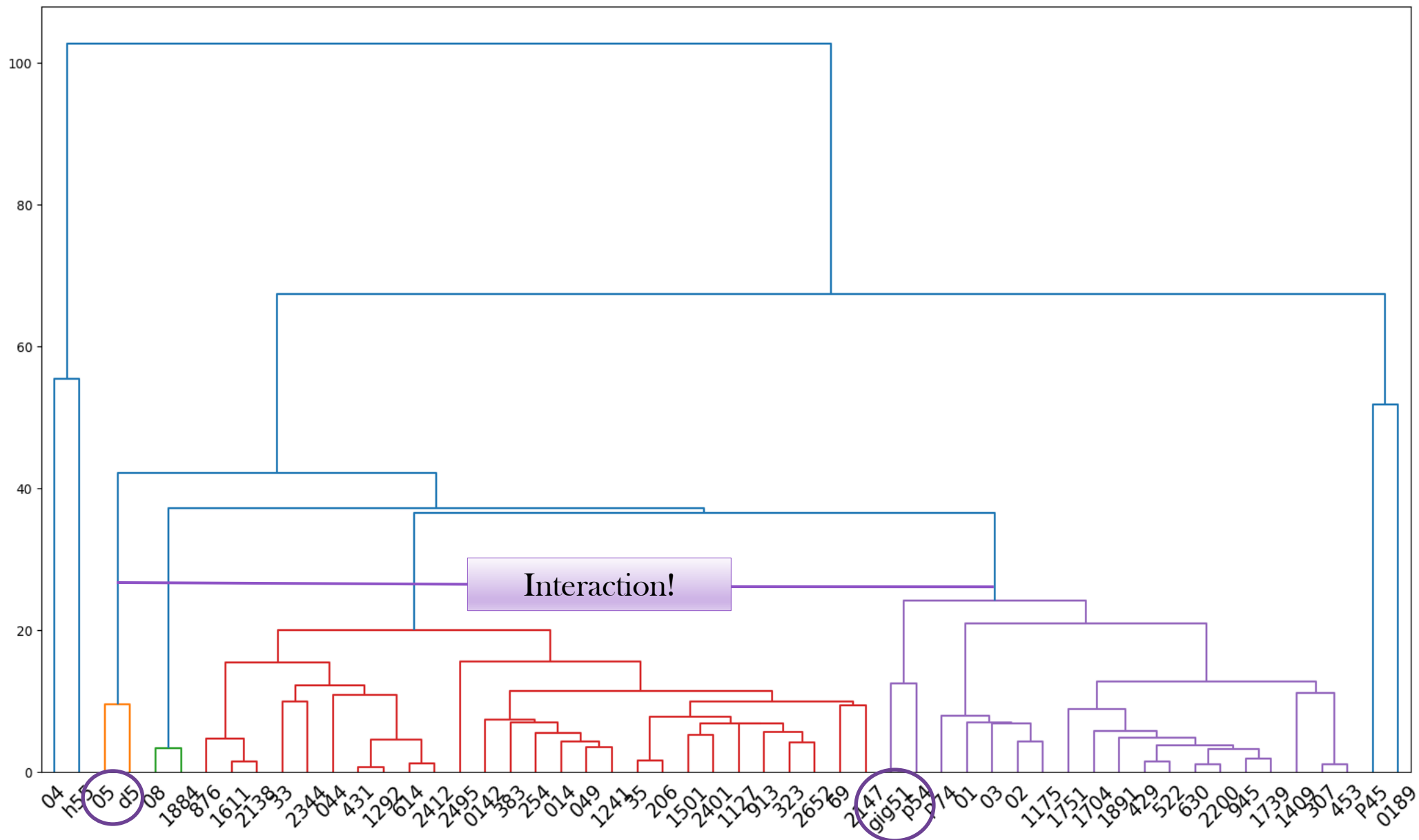


Second challenge

- Recombination
 - Scribes used more than one exemplar
 - Recombination mixes manuscript relations
- New Testament manuscripts are heavily recombined
 - Conventional hierarchical classifications becomes very difficult
 - Mixture blurs the boundaries of textual groups

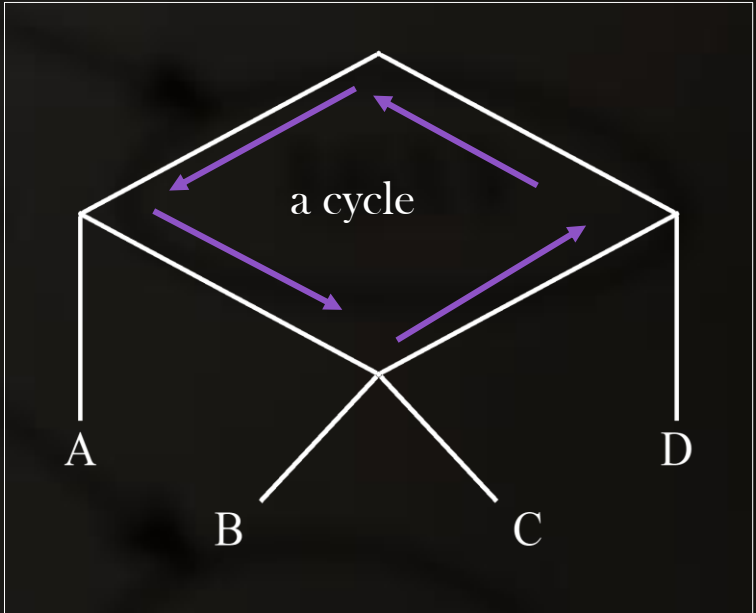
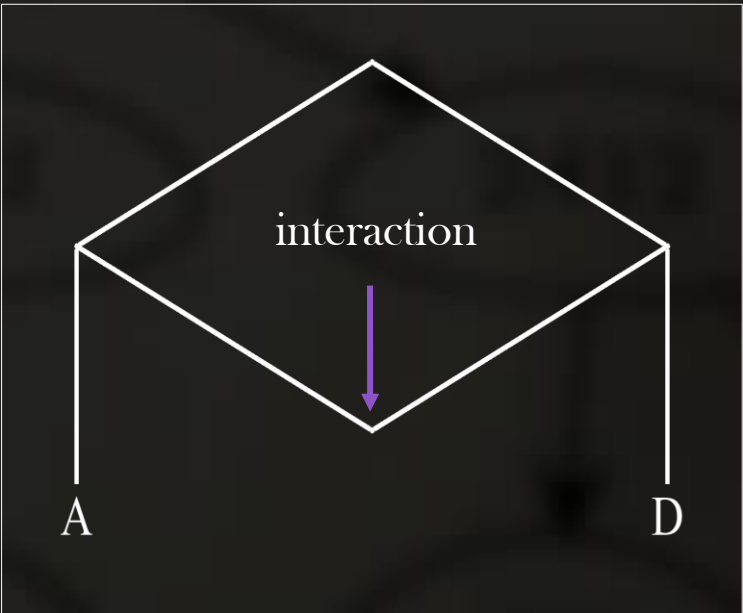
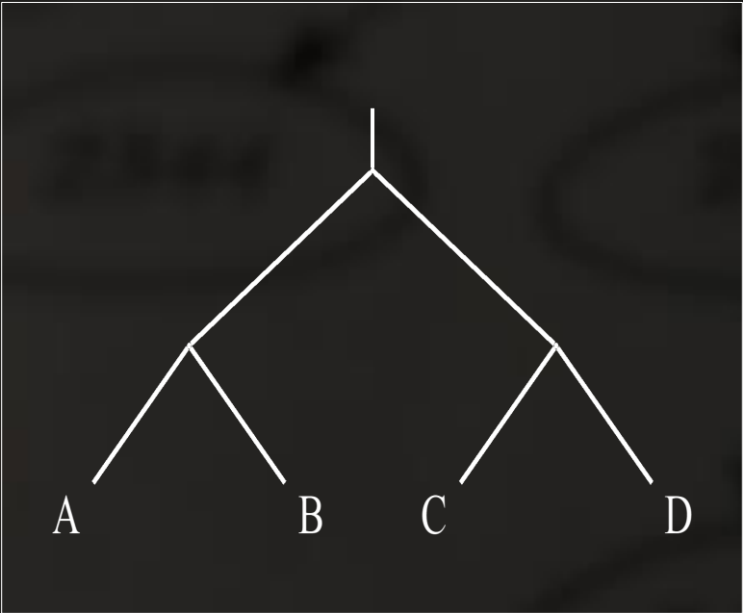








From trees to networks





Preprocessing the manuscript data for the network analysis

- Data mining techniques
 - The k-shingling algorithm
 - Texts are broken into smaller pieces (tokenization)
 - results to distance values
 - Every manuscript is compared with one another
 - A distance matrix



Preprocessing the manuscript data for the network analysis

- Data mining techniques

- The k

– Text

– resu

- Every

- A dist

Table 4. A distance matrix

	A	B	C	D	E	F
A	0	5	7	2	9	3
B	5	0	4	11	14	2
C	7	4	0	7	11	10
D	2	11	7	0	15	6
E	9	14	11	15	0	19
F	3	2	10	6	19	0



K-Shingling

a man saw that the fox jumped over the fence

a man

man saw

saw that

that the

the fox

fox jumped

jumped over

over the

the fence

that the

fox jumped

the fence

a man

jumped over

over the

the fox

saw that

man saw



K-Shingling

a man saw that the fox jumped over the fence

a man saw that the fox jumped over the hedge

fox jumped

man saw

jumped over

that the

saw that

the fox

a man

over the

the fence

that the

fox jumped

the hedge

a man

jumped over

over the

the fox

saw that

man saw



Calculating similarities

<i>Set1</i>	fox jumped	man saw	jumped over	that the	saw that	the fox	a man	over the	the fence
-------------	------------	---------	-------------	----------	----------	---------	-------	----------	-----------

<i>Set2</i>	that the	fox jumped	the hedge	a man	jumped over	over the	the fox	saw that	man saw
-------------	----------	------------	-----------	-------	-------------	----------	---------	----------	---------

Word bigram	<i>Set 1</i>	<i>Set 2</i>
fox jumped	1	1
man saw	1	1
jumped over	1	1
that the	1	1
saw that	1	1
the fox	1	1
a man	1	1
over the	1	1
the hedge	0	1
the fence	1	0



Intersection = 8

Union = 10

Sørensen-Dice Coefficient (SDC)

$$\frac{2 \times \text{intersection}}{\text{sum of the number of elements in each set}}$$

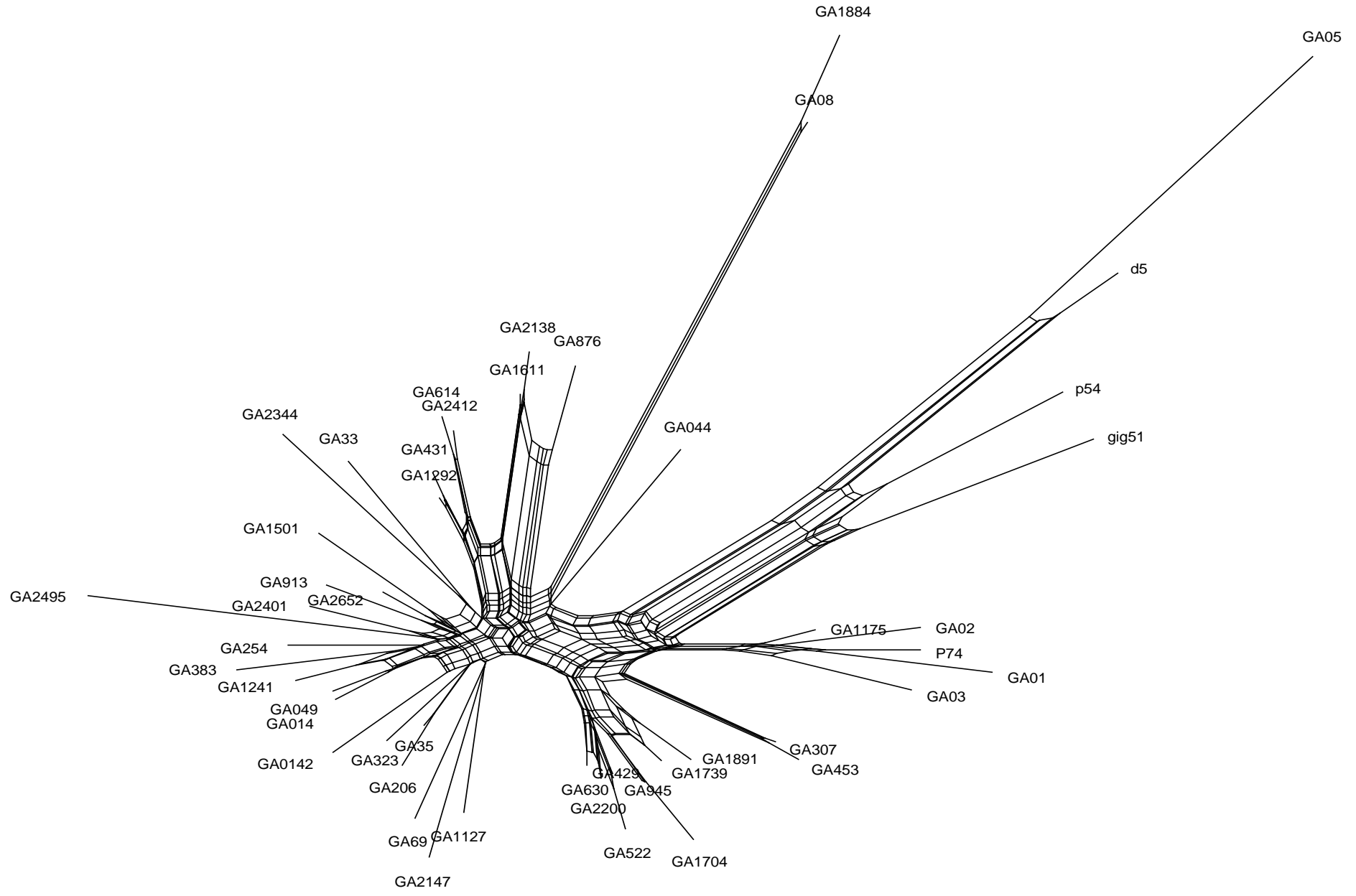
$$= 16 / 18 = 0.888 = 88 \%$$



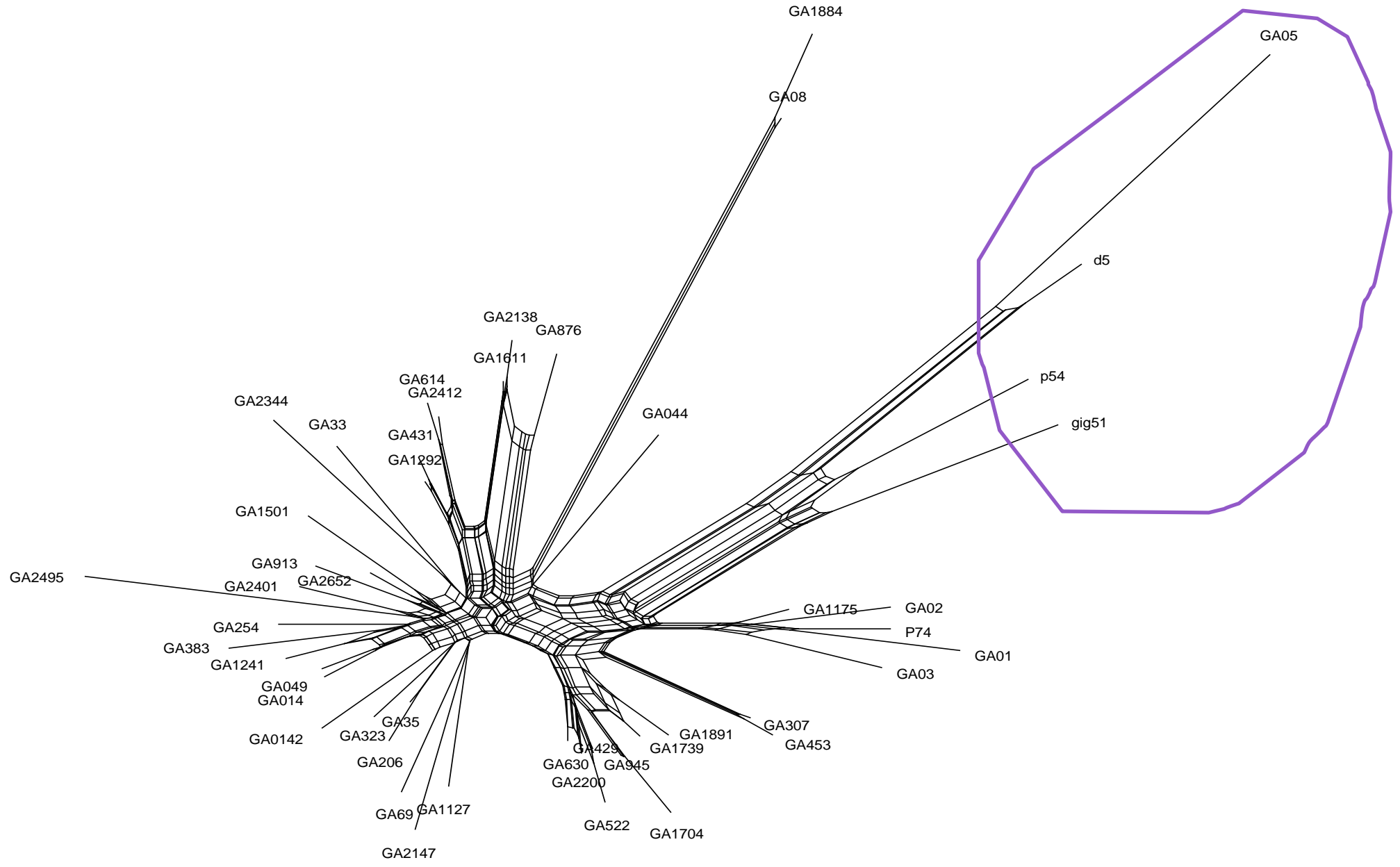
Preprocessing the manuscript data for network analysis

- The k-shingling algorithms can process hundreds of manuscripts in matter of minutes
- Distance values are used to construct a network

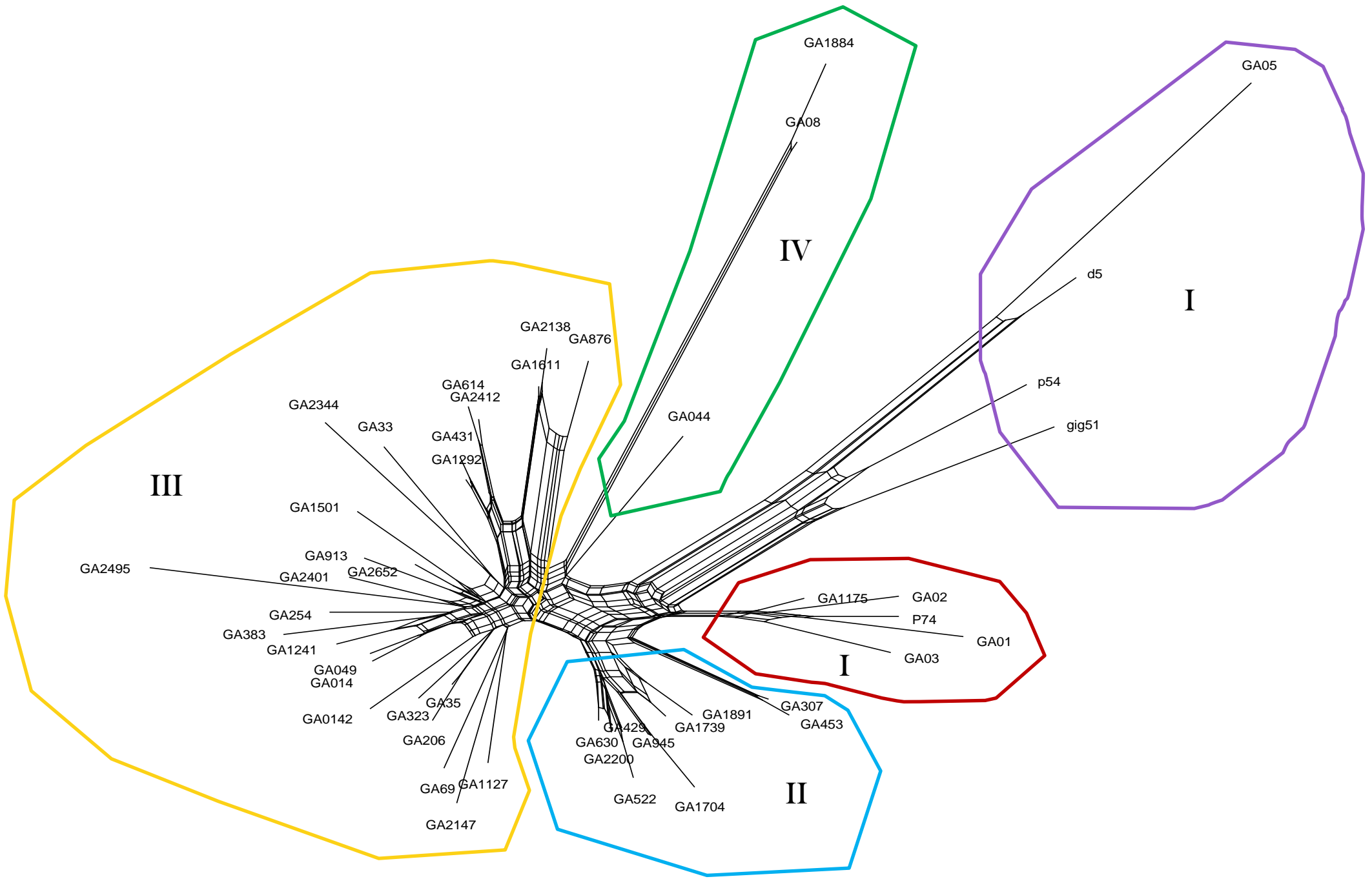
10.01



10.01



10.01





The future is in algorithms

- A Program called Relate was written to conduct the analysis
 - <https://github.com/PasHyde/relate>
- The possibilities of the network analysis are substantial
- The described method allows one to consider all manuscript evidence