# Network Analysis and the Classification of New Testament Manuscripts

Pasi Hyytiäinen

Faculty of Theology

*”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

# Manuscript traditions of the New Testament

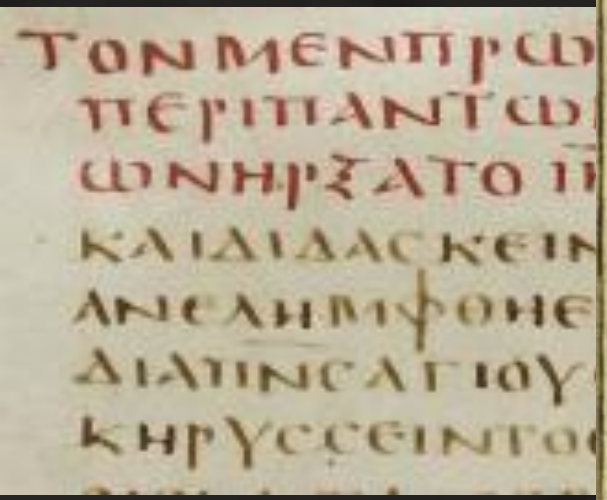
- New Testament was transmitted through the medium of manuscripts for over 1400 years
  - Manuscript = manus scriptus (“written by hand”)

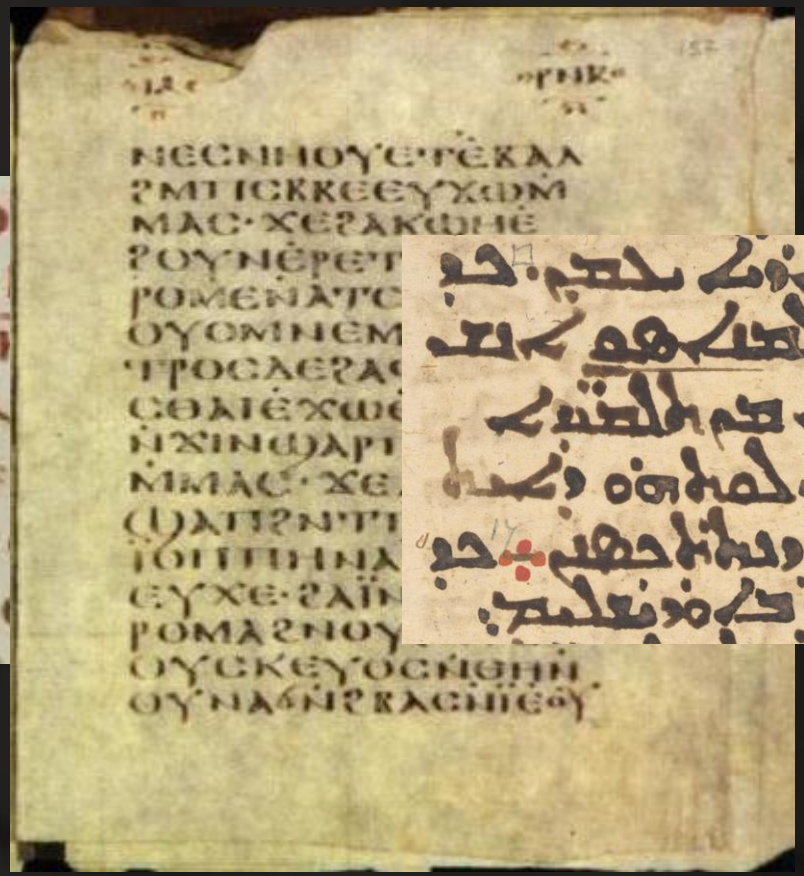
Source: British Library Catalogue of Illuminated Manuscripts. London, British Library MS Royal 14 E III, f. 6v. Northern France, 1st quarter of the 14th century
http://www.bl.uk/catalogues/illuminatedmanuscripts/ILLUMIN.ASP?Size=mid&IllID=43454

# Manuscript traditions of the New Testament

- 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







# First Challenge

- The sizes of the manuscript traditions
  - Approximately 6000 Greek
  - 10 000 Latin
  - 1500 Armenian
  - 1000 Coptic New Testament manuscripts
- New manuscript discoveries are made each year

WWU MÜNSTER

[INTF] INSTITUT FÜR NEUTESTAMENTLICHE TEXTFORSCHUNG

Sign In

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 - ∞

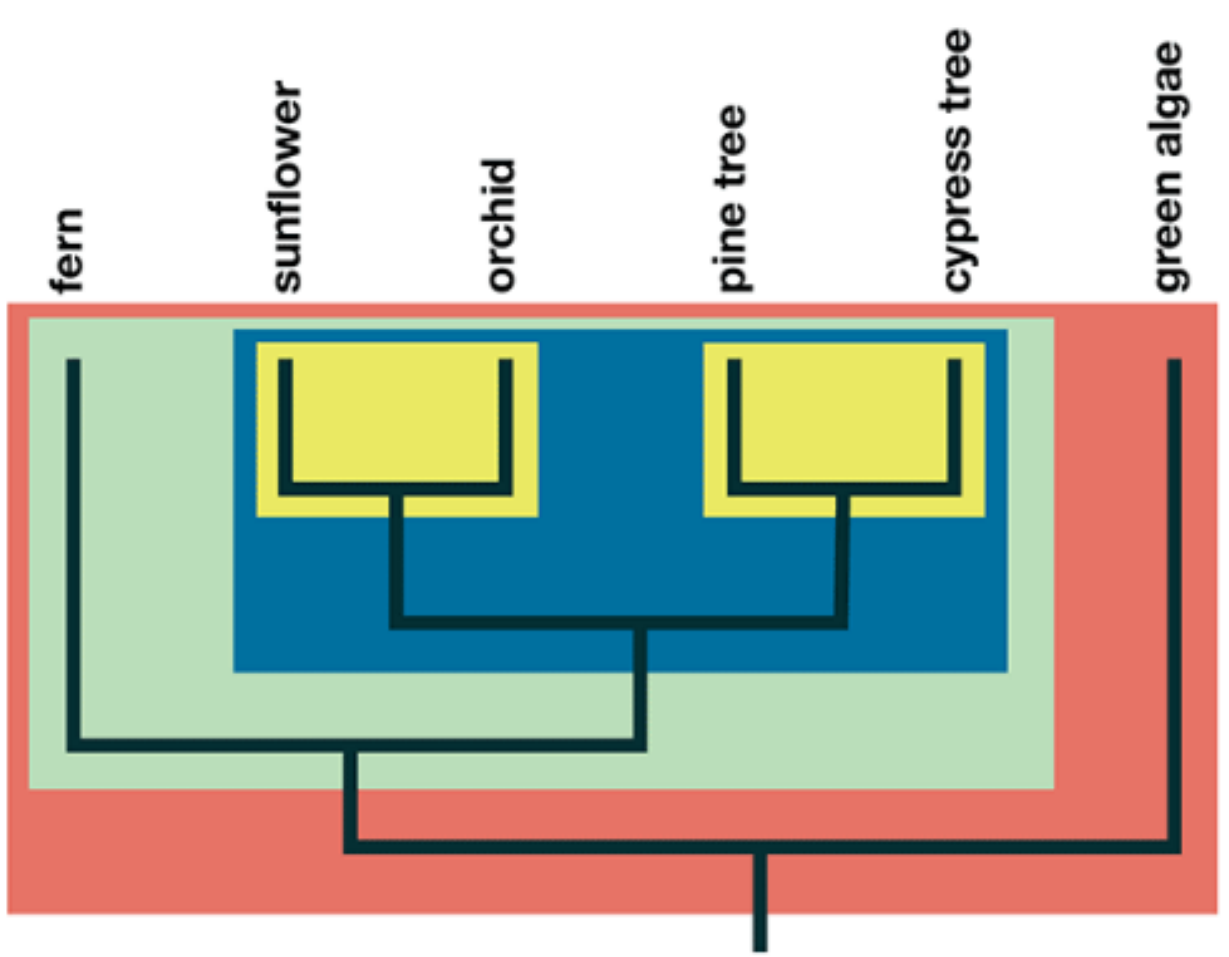
Has: Images Transcriptions

## Manuscript Details

## Bibliography

# First Challenge

- 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




ze the data

rld in a similar

# First Challenge

- Considering all manuscript data is impossible when conventional computer-assisted 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

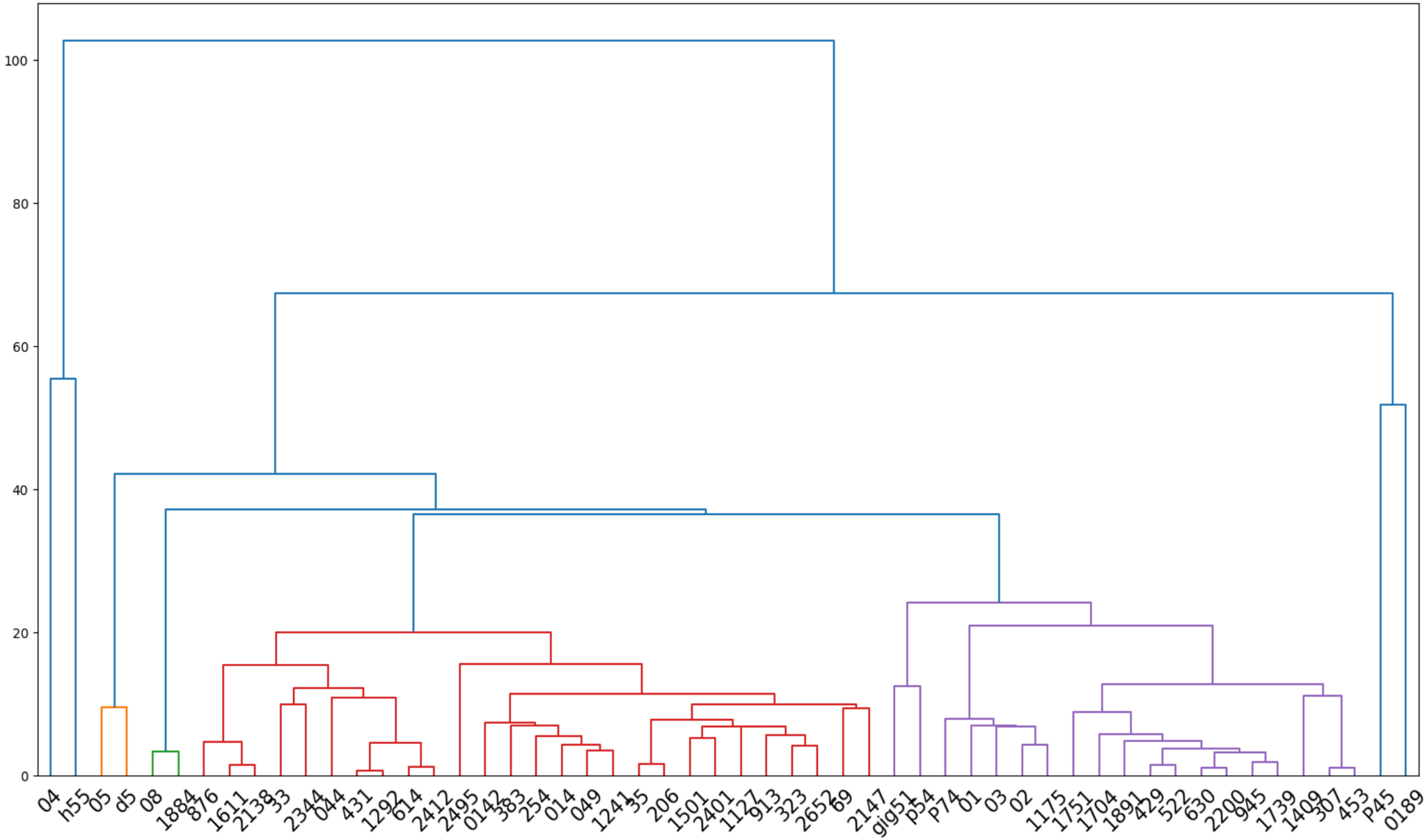
| | Solution 1 | | Solution 2 | | |
|---|---|---|---|---|---|
| A | – | the fox jumped over the hedge | – | the fox jumped | over the hedge |
| B | – | – | – | – | – |
| C | – | the cat jumped over the fence | – | the cat jumped | over the fence |
| D | a man saw that | the fox jumped over the hedge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

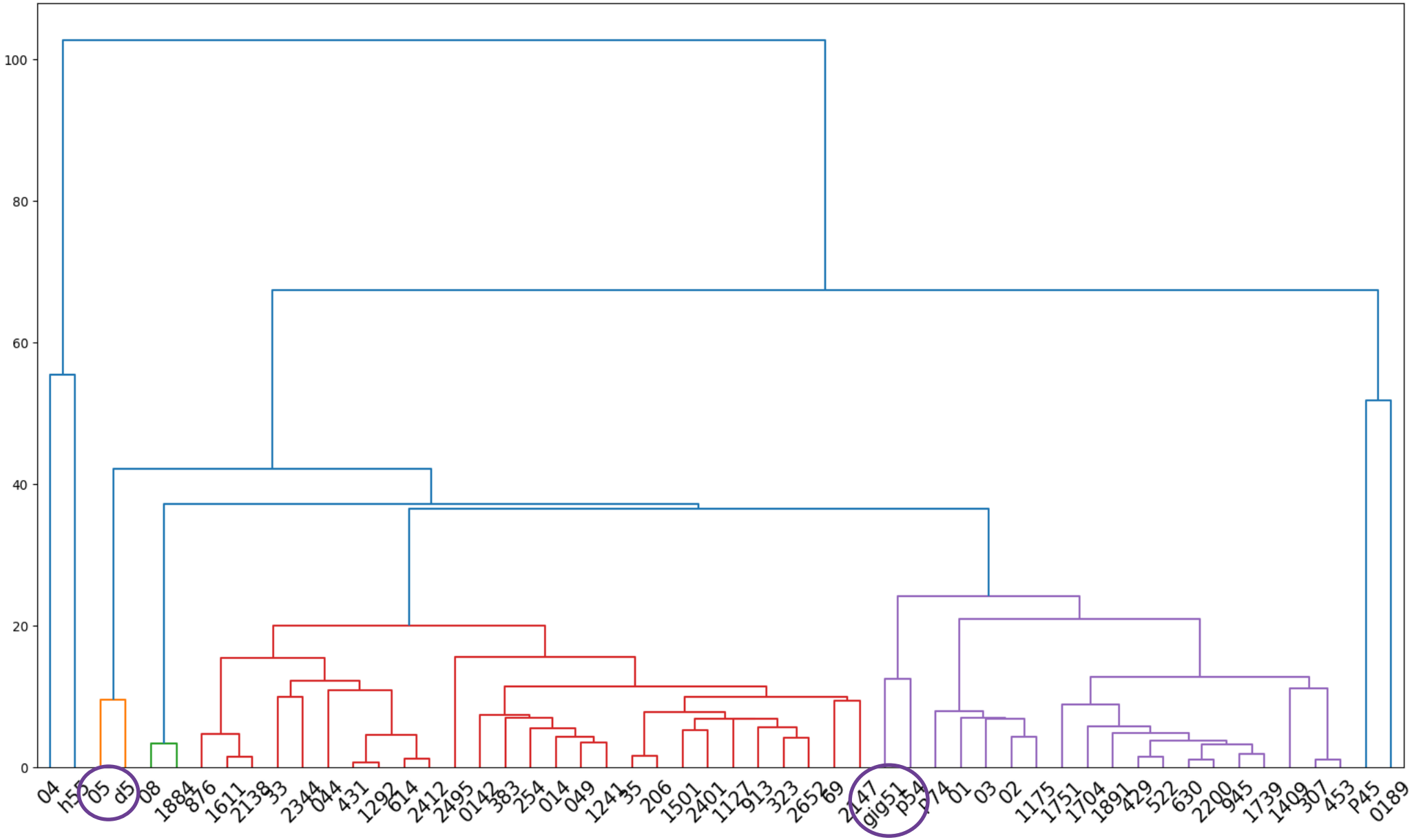
# Encoding variation places

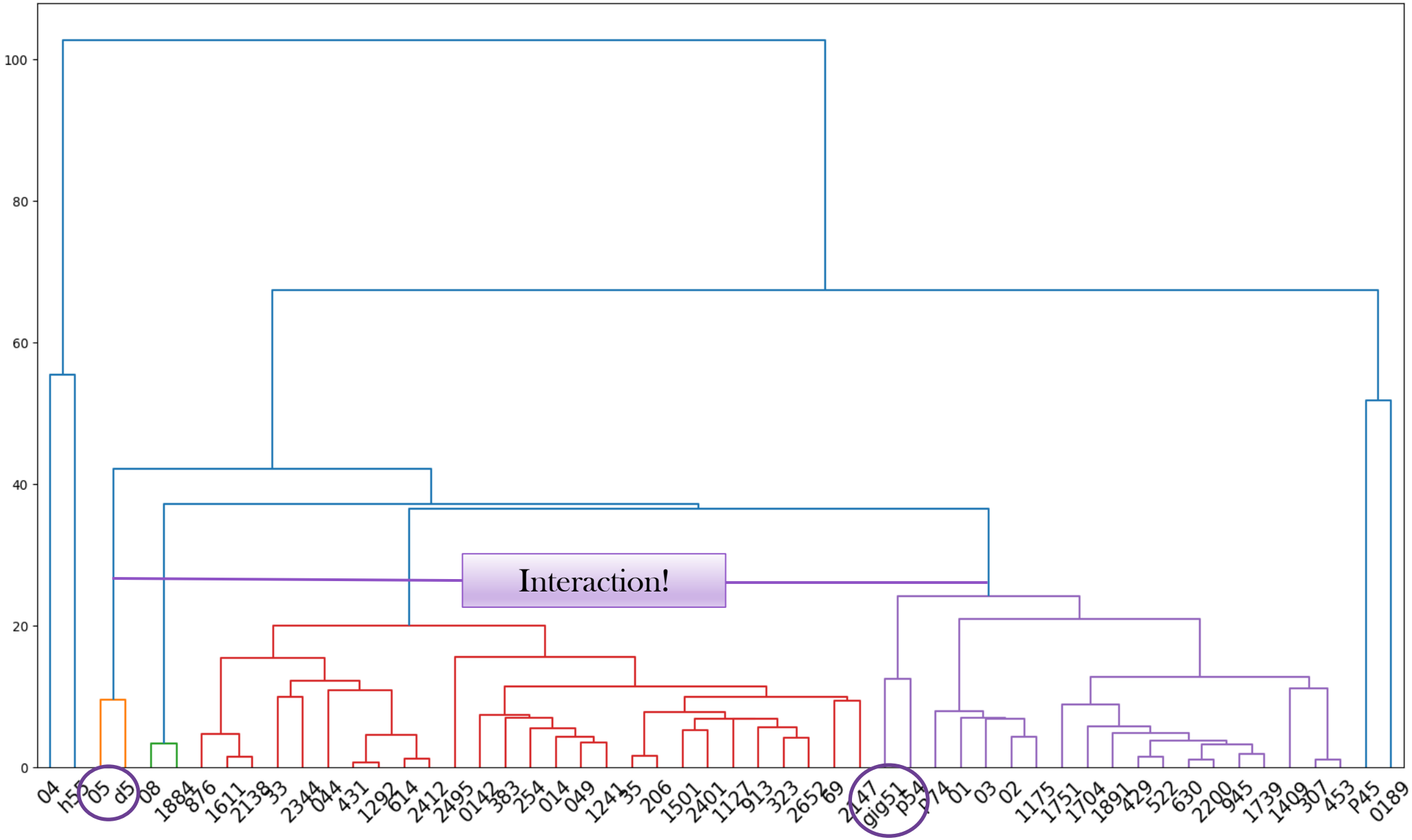
| | Solution 1 | | Solution 2 | | |
|---|---|---|---|---|---|
| A | 0 | 1 | 0 | 0 | 0 |
| B | ? | ? | ? | ? | ? |
| C | 0 | 0 | 0 | 1 | 1 |
| D | 1 | 1 | 1 | 0 | 0 |
| E | 1 | 2 | 1 | 0 | 1 |

# Second challenge

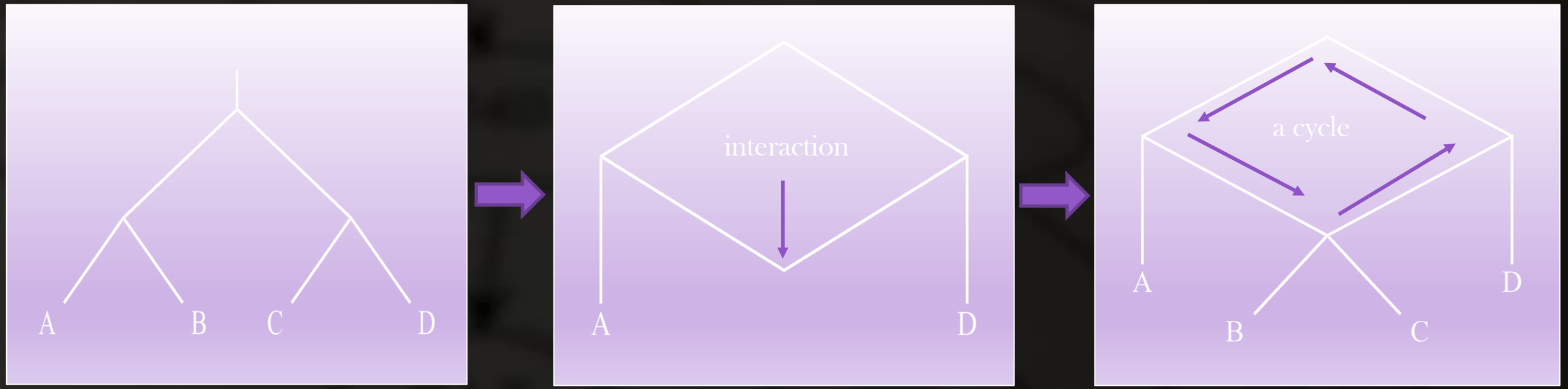
- Recombination
  - Scribes used more than one exemplar
  - Recombination mixes manuscript relations
- New Testament manuscripts are heavily recombined
  - Conventional hierarchical classifications becomes very difficult
  - Mixture blurs the boundaries of textual groups


100
80
60
40
20
0
04
h55
05
d5
08
1884
876
1611
2138
33
2344
044
431
1292
614
2412
2495
0142
383
254
014
049
1241
35
206
1501
2401
1127
913
323
2652
69
2147
gig51
p54
P74
01
03
02
1175
1751
1704
1891
429
522
630
2200
945
1739
1409
307
453
P45
0189


100
80
60
40
20
0
04
h55
05
d5
08
1884
876
1611
2138
33
2344
044
431
1292
614
2412
2495
0142
383
254
014
049
1241
35
206
1501
2401
1127
913
323
2652
69
2447
gig51
p54
p74
01
03
02
1175
1751
1704
1891
429
522
630
2200
945
1739
1409
307
453
P45
0189


100
80
60
40
20
0
Interaction!
04
h55
05
d5
08
1884
876
1611
2138
33
2344
044
431
1292
614
2412
2495
0142
383
254
014
049
1241
35
206
1501
2401
1127
913
323
2652
69
2447
gig51
p54
p74
01
03
02
1175
1751
1704
1891
429
522
630
2200
945
1739
1409
307
453
P45
0189

# 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
    - Tex
    - resu
  - Every
  - A dist

n)

r

*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

| *Set1* | fox jumped | man saw | jumped over | that the | saw that | the fox | a man | over the | the fence |
|---|---|---|---|---|---|---|---|---|---|

| *Set2* | that the | fox jumped | the hedge | a man | jumped over | over the | the fox | saw that | man saw |
|---|---|---|---|---|---|---|---|---|---|

| Word bigram | *Set 1* | *Set 2* |
|---|---|---|
| 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

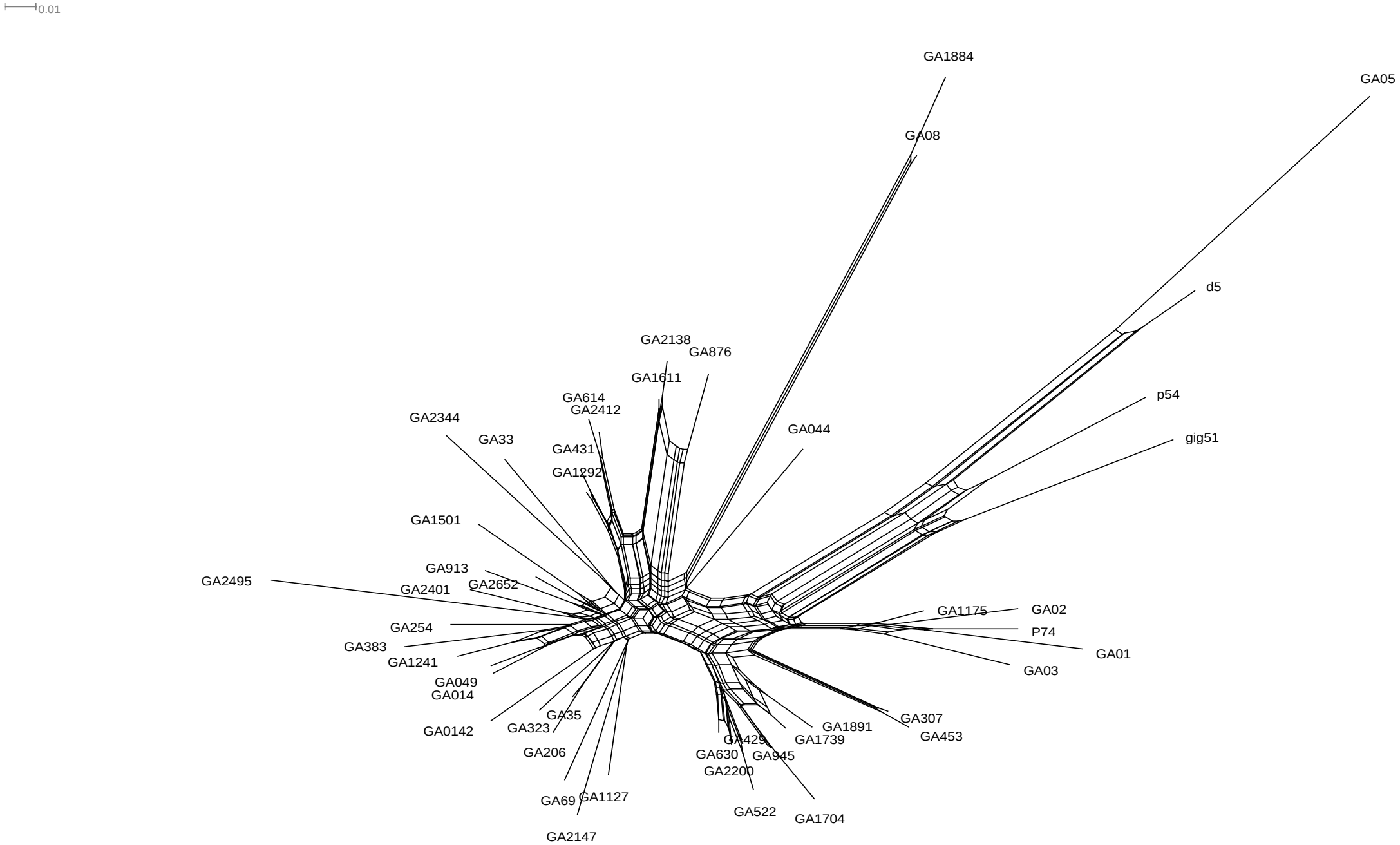
Sørensen-Dice Coefficient (SDC)

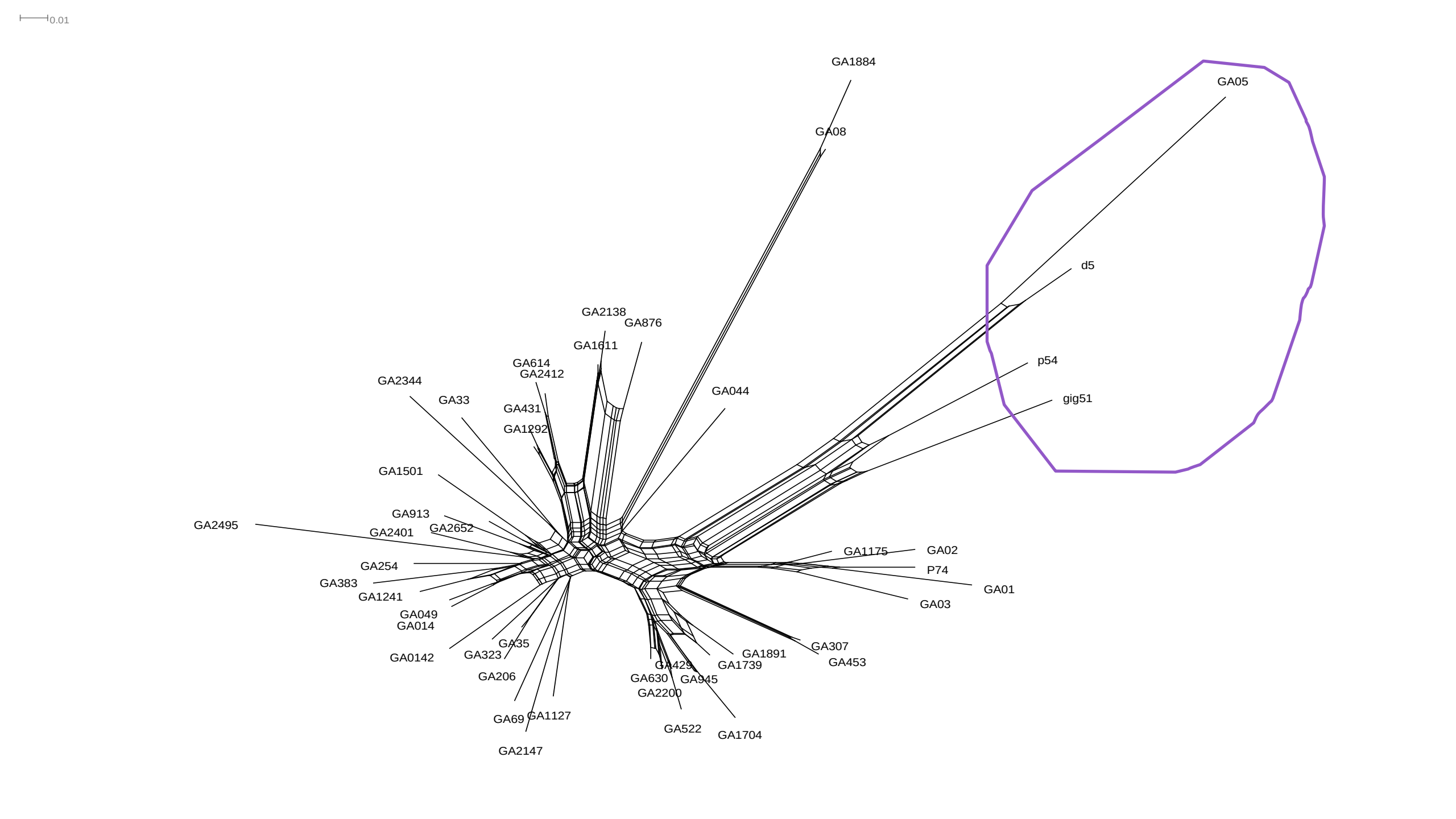
$$\frac{2\,x\,intersection}{sum\ of\ the\ number\ of\ elements\ in\ each\ set}$$

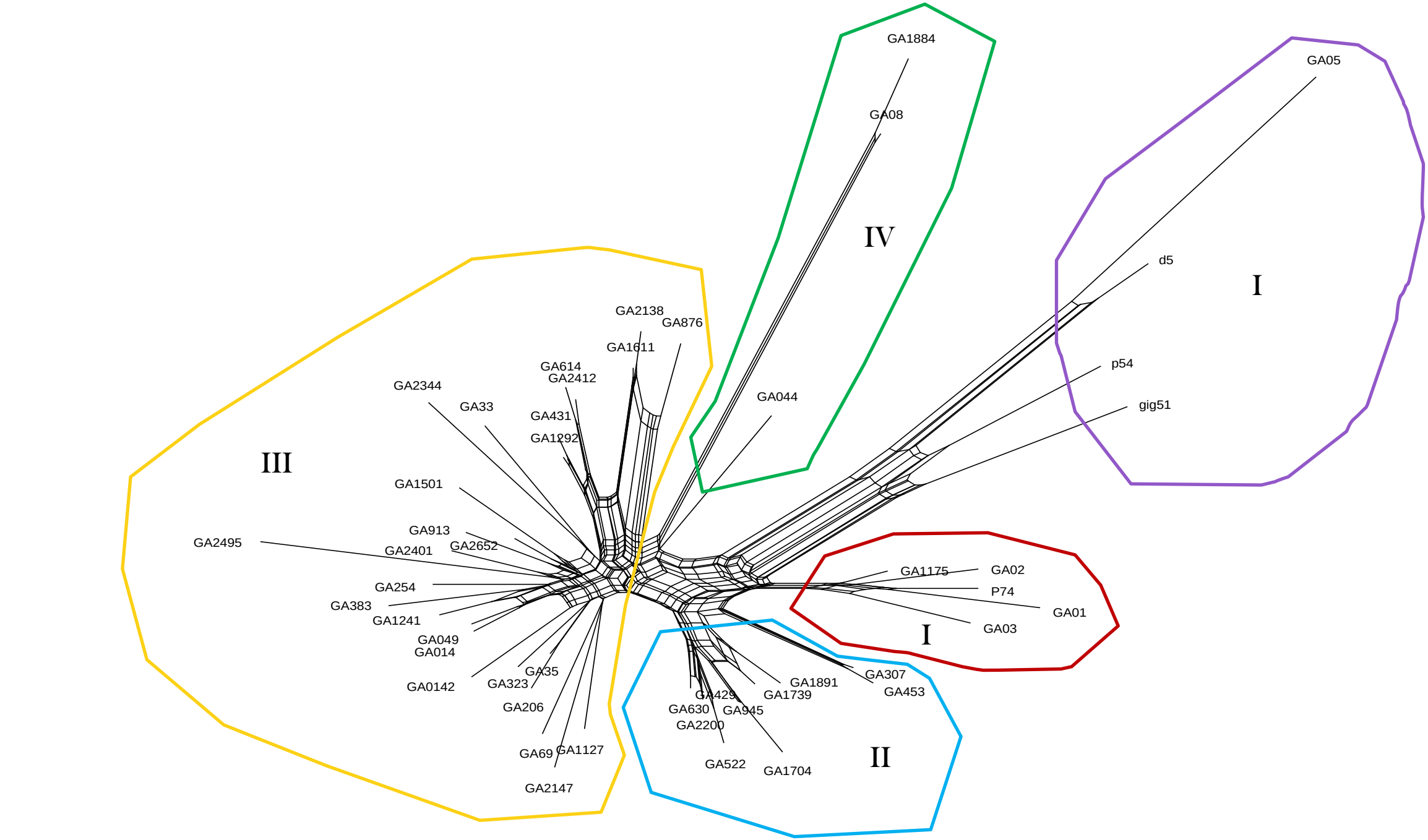
= 16 / 18 = 0.888 = 88 %

# 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


0.01
GA1884
GA05
GA08
d5
GA2138
GA876
GA1611
GA614
GA2412
p54
GA2344
GA33
GA044
gig51
GA431
GA1292
GA1501
GA913
GA2495
GA2652
GA2401
GA1175
GA02
GA254
P74
GA383
GA01
GA1241
GA03
GA049
GA014
GA307
GA35
GA1891
GA0142
GA323
GA453
GA1739
GA206
GA630
GA945
GA2200
GA69
GA1127
GA522
GA1704
GA2147


0.01
GA1884
GA08
GA05
d5
p54
gig51
GA2138
GA876
GA1611
GA614
GA2412
GA2344
GA33
GA431
GA1292
GA044
GA1501
GA913
GA2495
GA2652
GA2401
GA254
GA383
GA1241
GA049
GA014
GA0142
GA323
GA35
GA206
GA69
GA1127
GA2147
GA1175
GA02
P74
GA01
GA03
GA307
GA453
GA1891
GA1739
GA429
GA630
GA945
GA2200
GA522
GA1704


0.01
GA1884
GA08
IV
GA05
d5
I
p54
gig51
GA044
GA2138
GA876
GA1611
GA614
GA2412
GA2344
GA33
GA431
GA1292
III
GA1501
GA913
GA2495
GA2652
GA2401
GA254
GA383
GA1241
GA049
GA014
GA0142
GA323
GA35
GA206
GA69
GA1127
GA2147
GA1175
GA02
P74
GA01
GA03
I
GA307
GA1891
GA453
GA1739
GA429
GA630
GA945
GA2200
GA522
GA1704
II

# 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