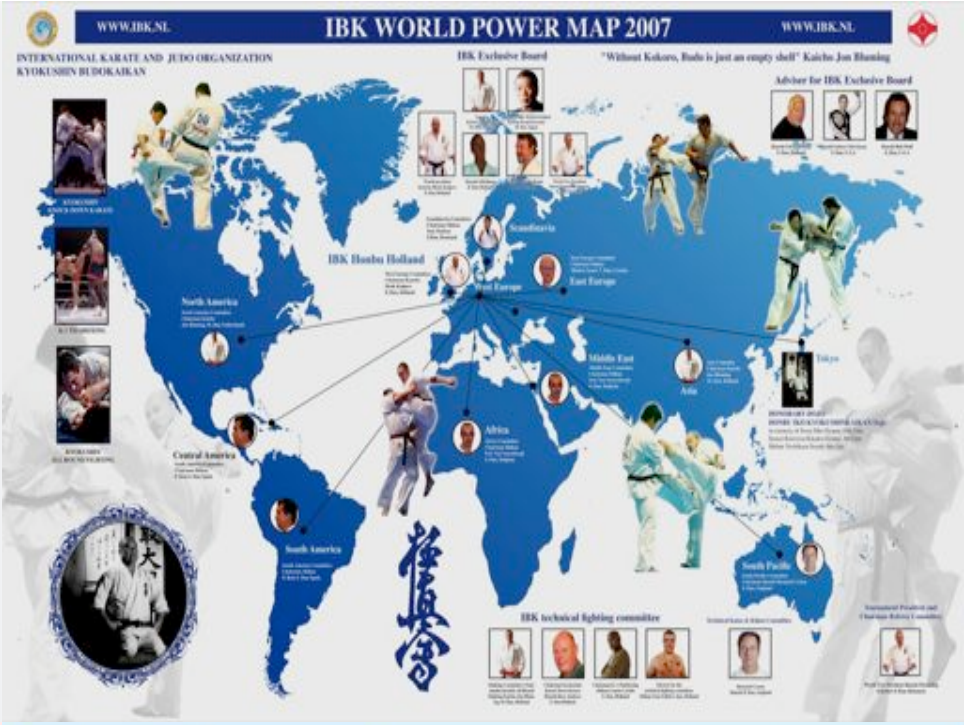
Consistent Layout for Thematic Software Maps

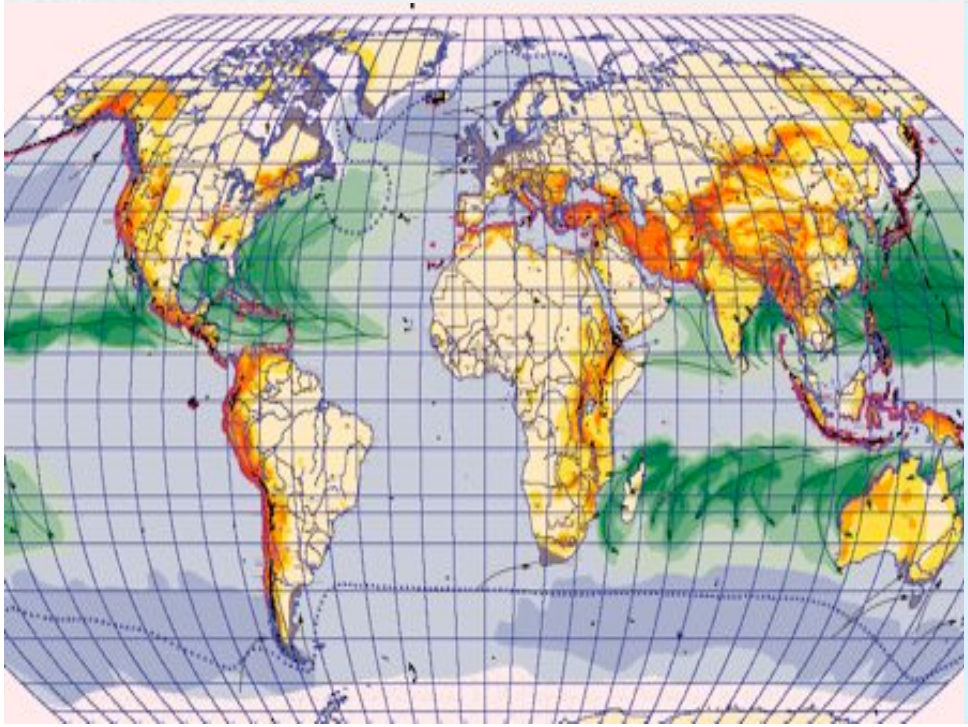
Adrian Kuhn, Peter Loretan, **Oscar Nierstrasz**

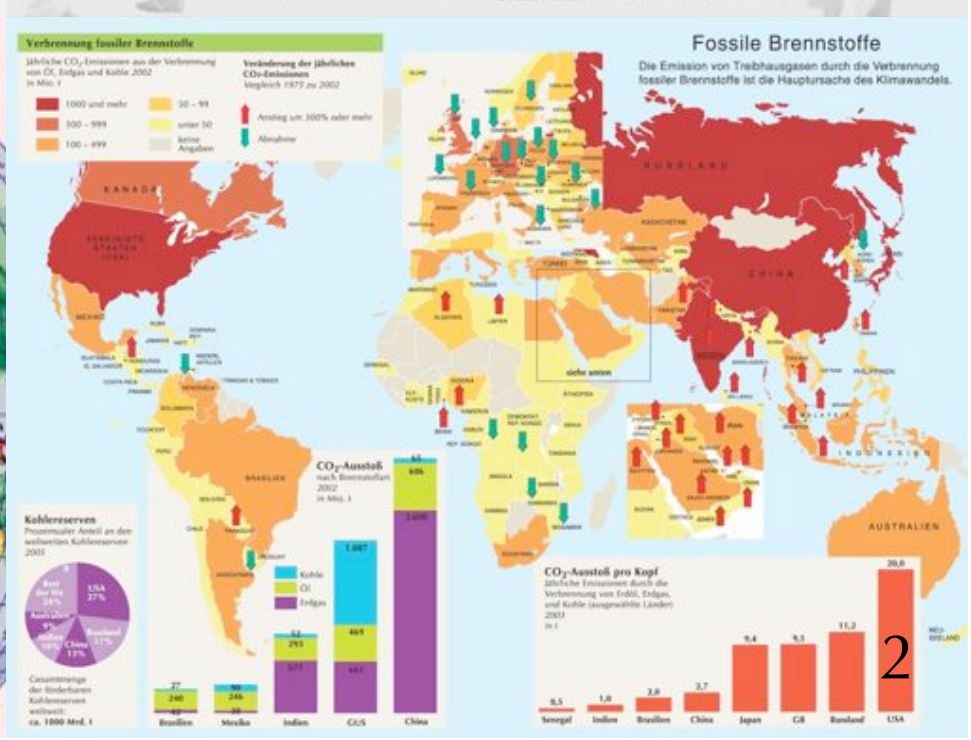
SCG, University of Bern

Published and presented at WCRE 2008.


Globale Handelsströme

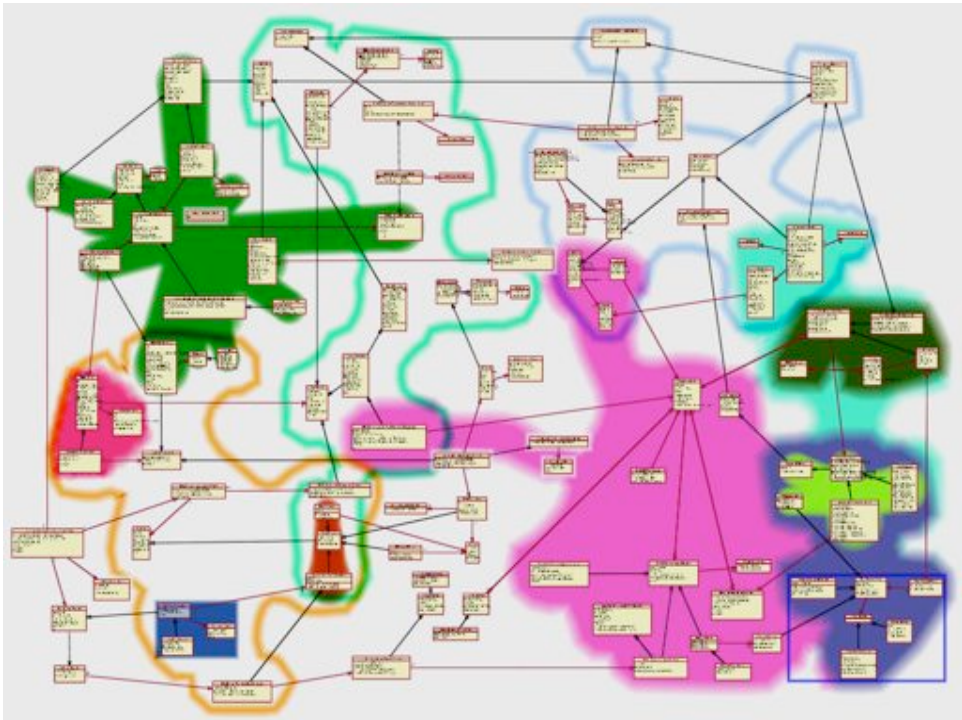

WWW.IBK.NL
IBK WORLD POWER MAP 2007
WWW.IBK.NL
INTERNATIONAL KARATE AND JUDO ORGANIZATION
KYOKUSHIN BUDOKAIKAN
IBK Exclusive Board
"Without Kokoro, Budo is just an empty shell" Kaicho Jon Bluming
Adviser for IBK Exclusive Board
IBK Honbu Holland
North America
Central America
South America
Africa
Middle East
Asia
South Pacific
Tokyo
IBK technical fighting committee

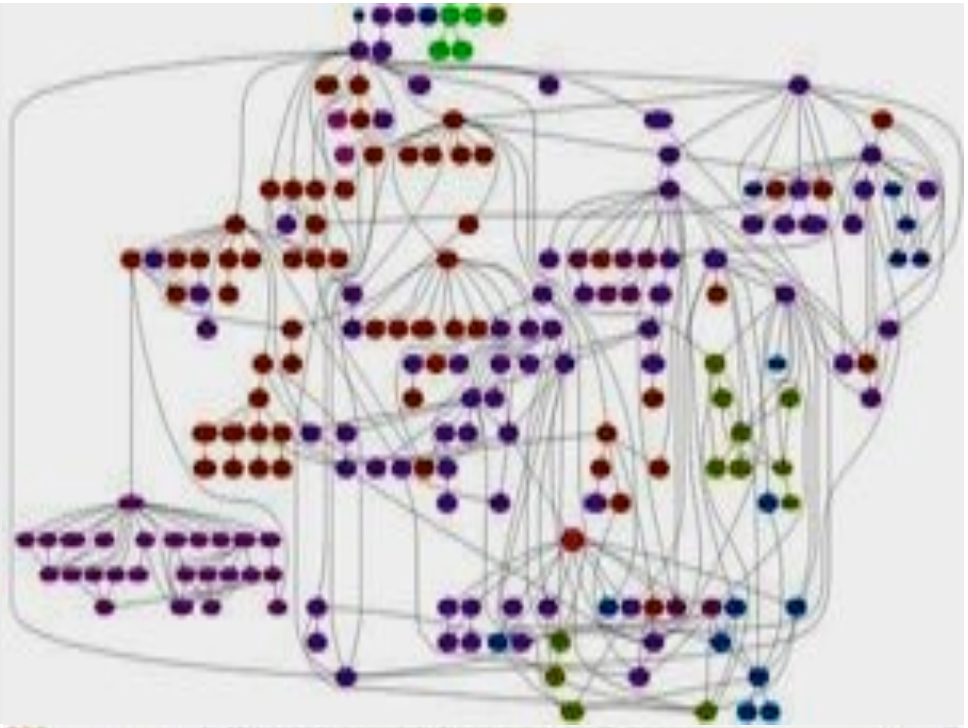


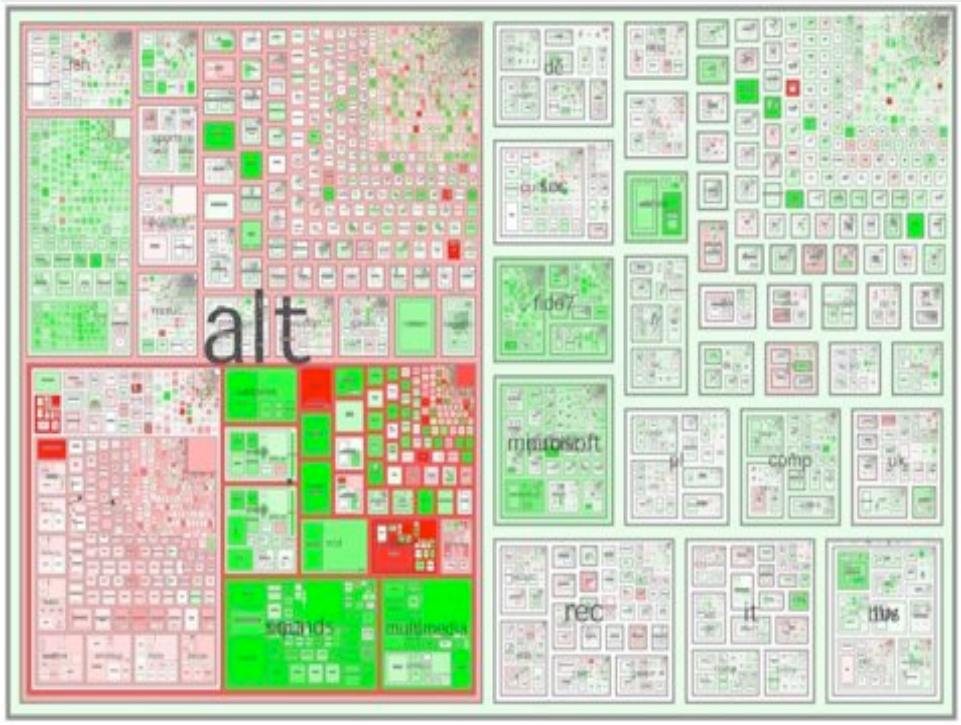

Verbrennung fossiler Brennstoffe
Fossile Brennstoffe
Die Emission von Treibhausgasen durch die Verbrennung fossiler Brennstoffe ist die Hauptursache des Klimawandels.
Kohlereserven
CO2-Ausstoß
CO2-Ausstoß pro Kopf

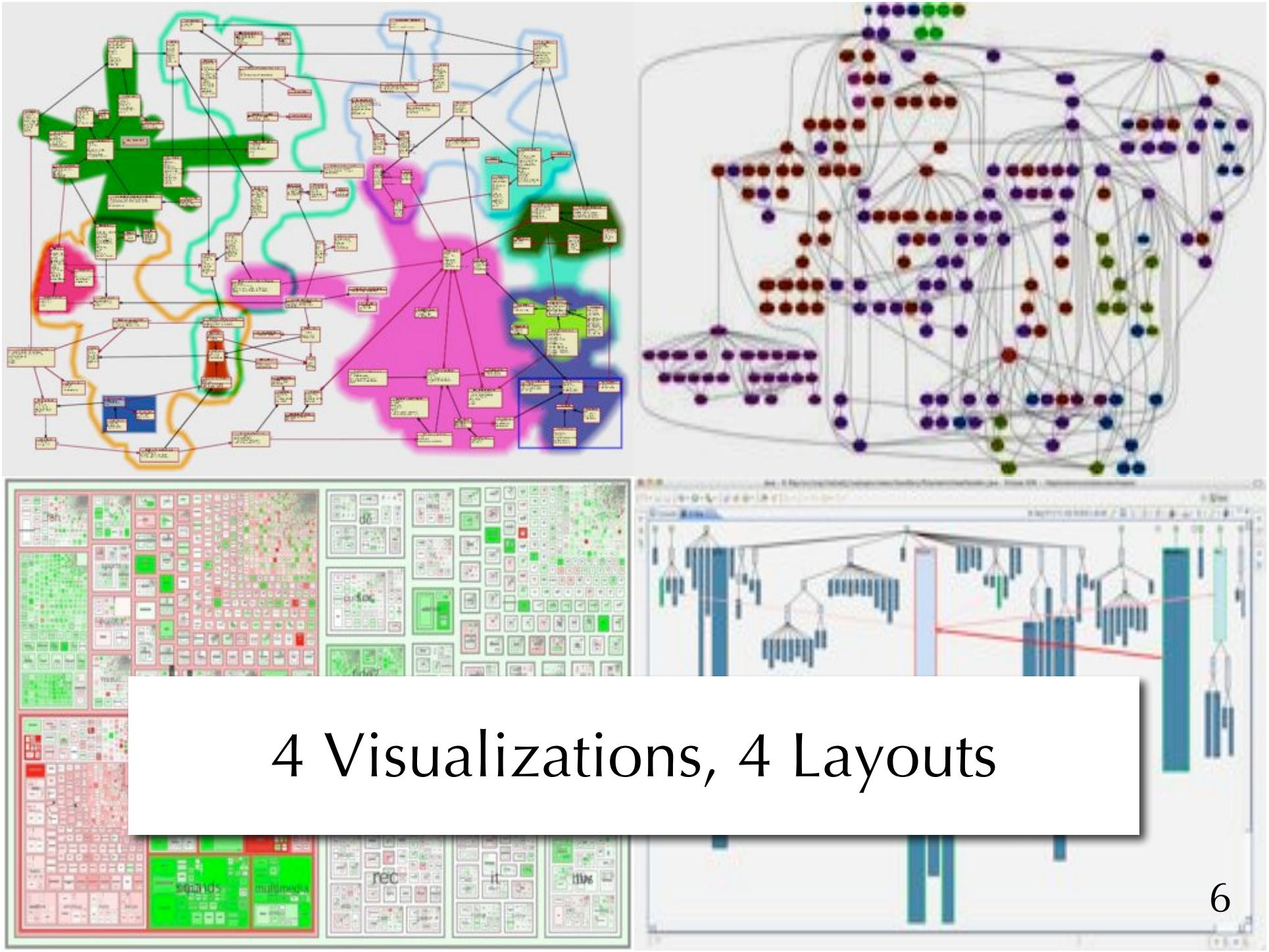

Globale Handelsströme
IBK WORLD POWER MAP 2007
WWW.IBK.NL
Verbrennung fossiler Brennstoffe
Fossile Brennstoffe
Die Emission von Treibhausgasen durch die Verbrennung fossiler Brennstoffe ist die Hauptursache des Klimawandels.
4 Visualizations, 1 Layout

Distance has a meaning

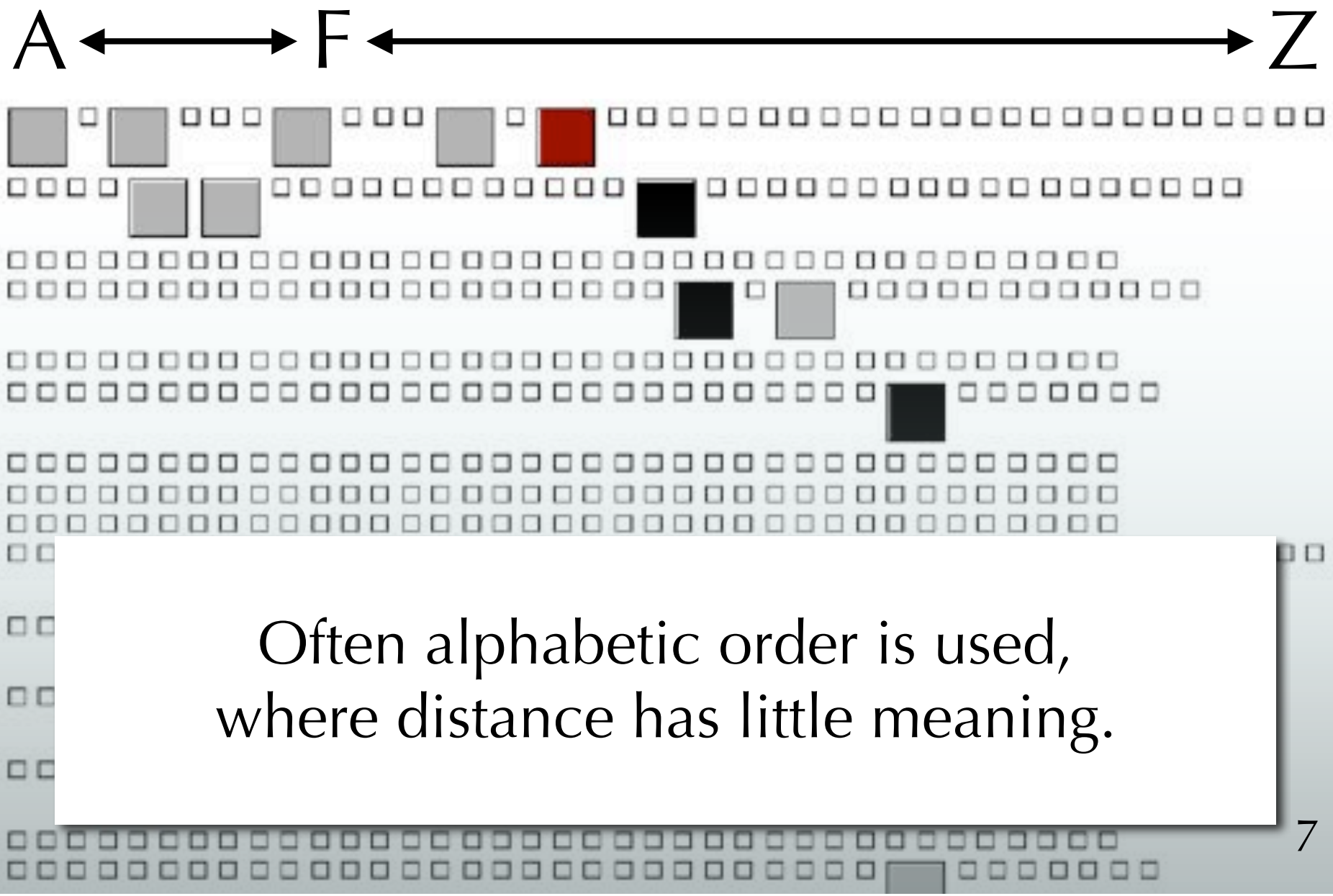





alt
rec

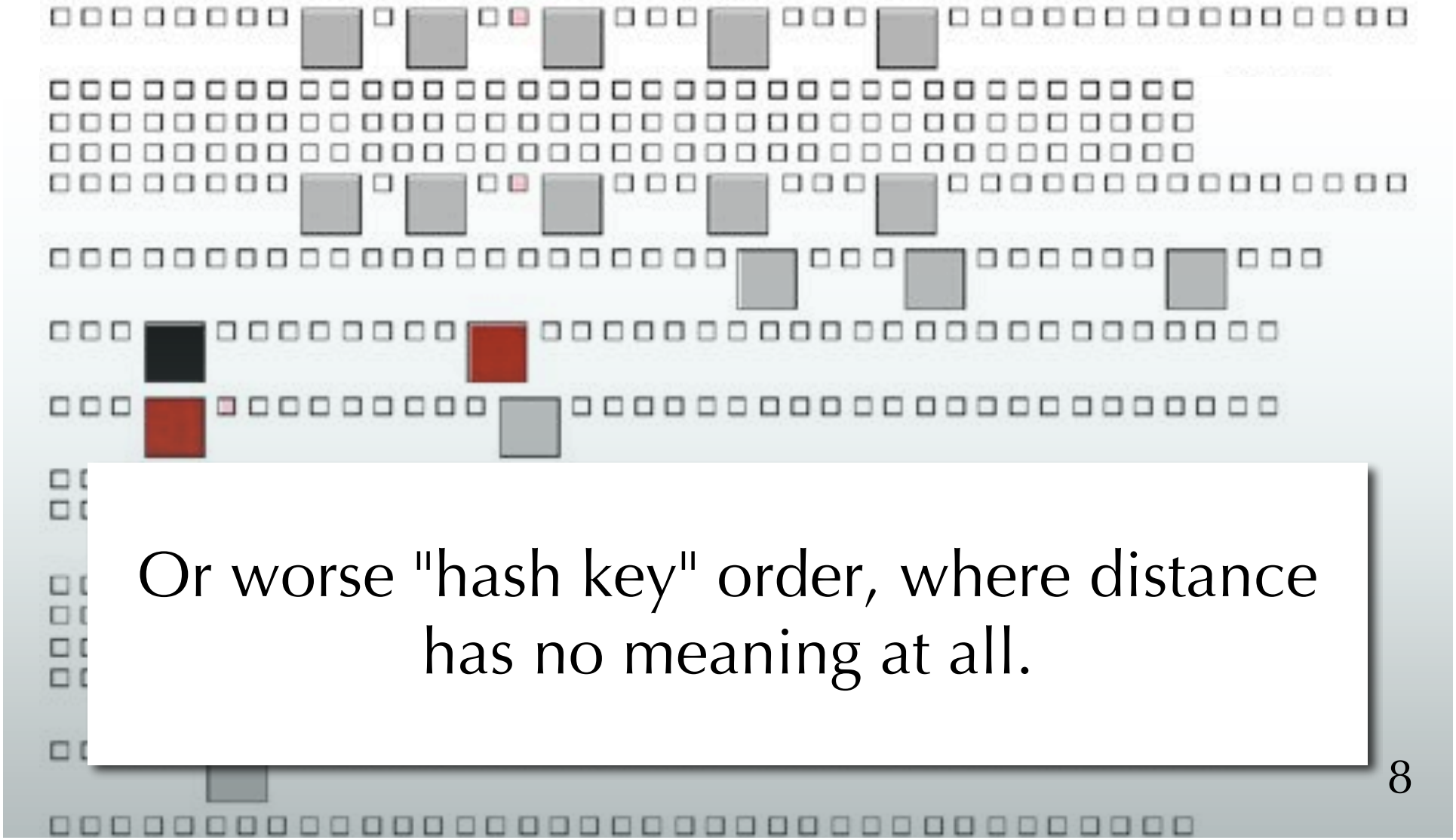


4 Visualizations, 4 Layouts



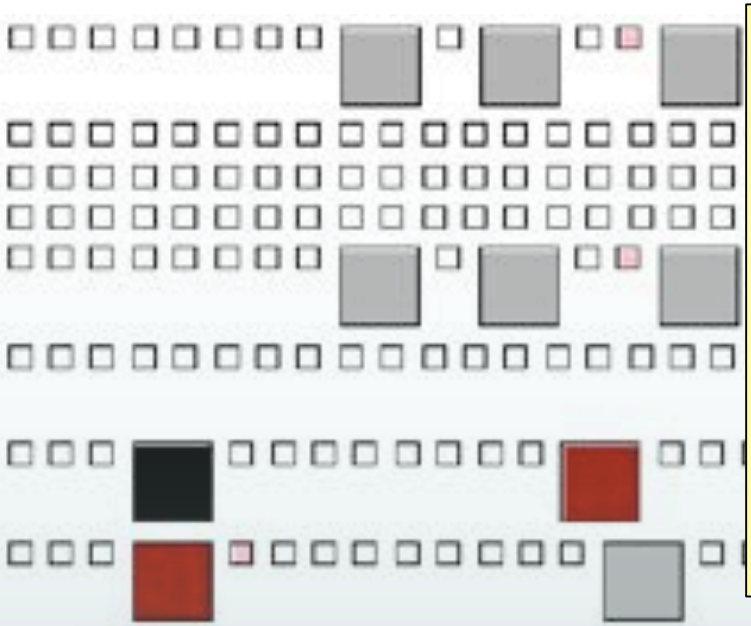
Often alphabetic order is used,
where distance has little meaning.

0x017 ⟷ 0x02A ⟷ 0x7FF



Or worse "hash key" order, where distance has no meaning at all.

0x017 ⟷ 0x02A ⟷ 0x7FF



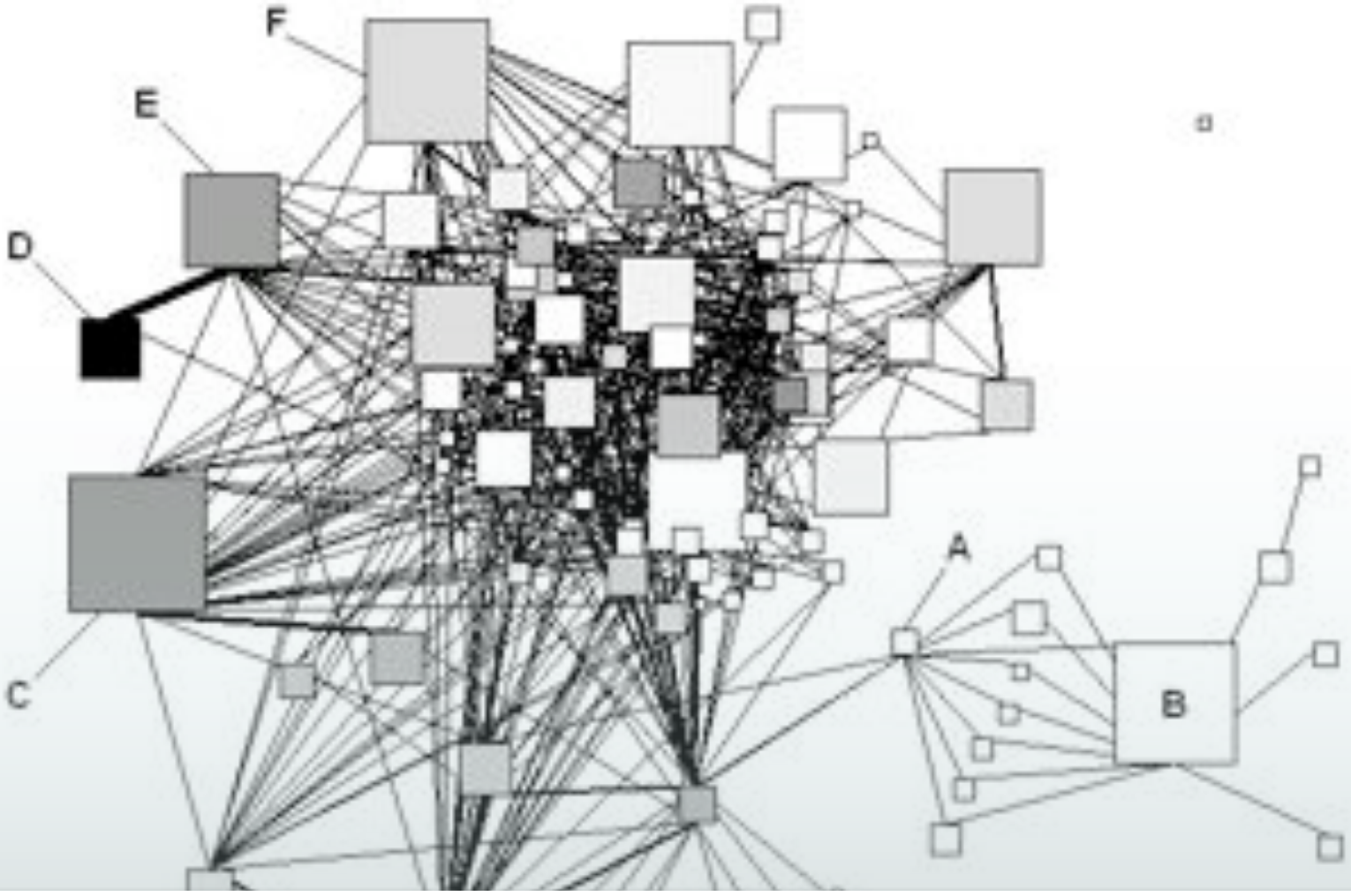
```
for (each : Set all) {
  // non-deterministic order!
}
```

Why? Set is internally realized as a dictionary. Iteration order is given by hash keys, which are arbitrary and thus meaningless.

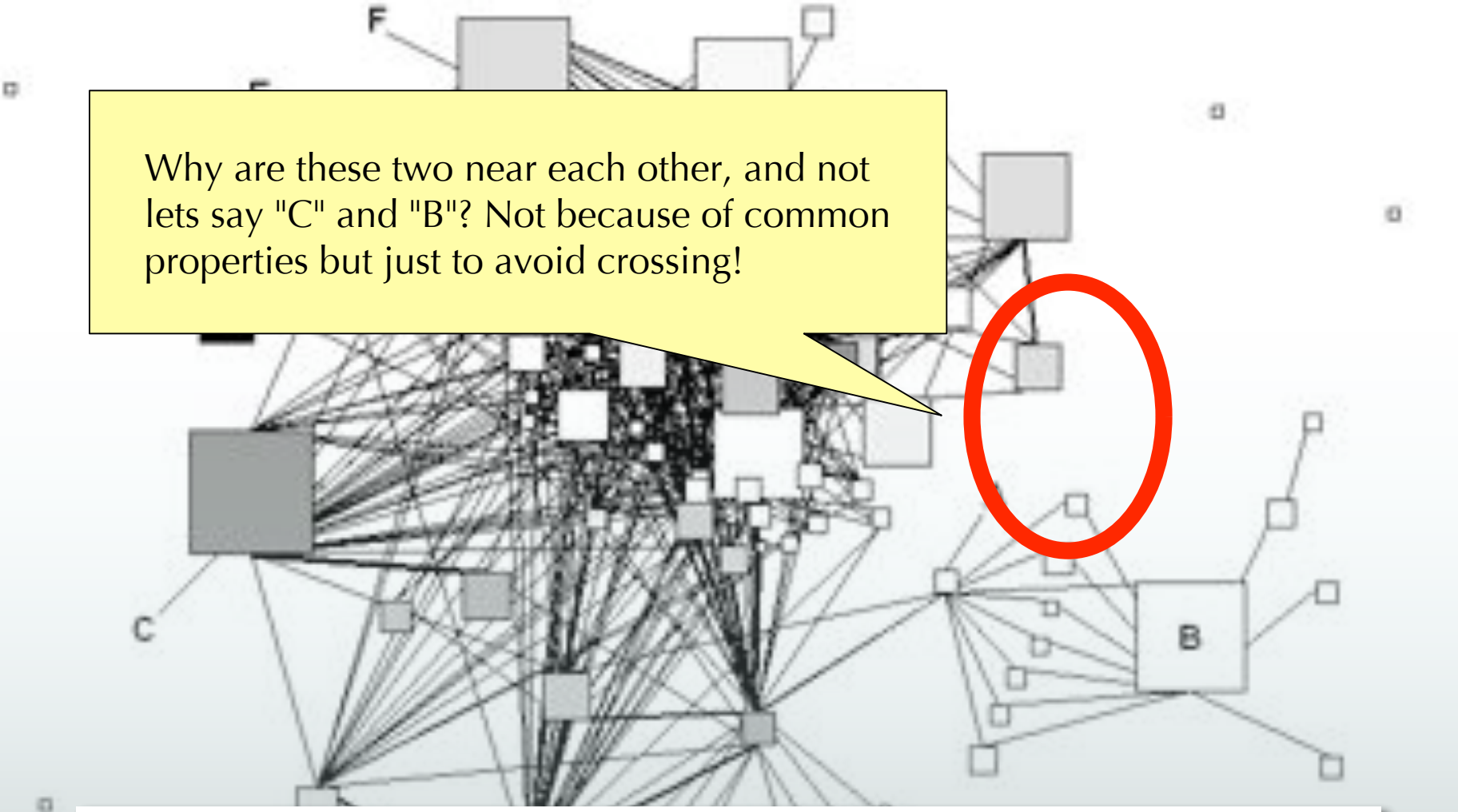
Or worse "hash key" order, where distance has no meaning at all.

We need: a consistent and meaningful layout for software.

Has it been done?

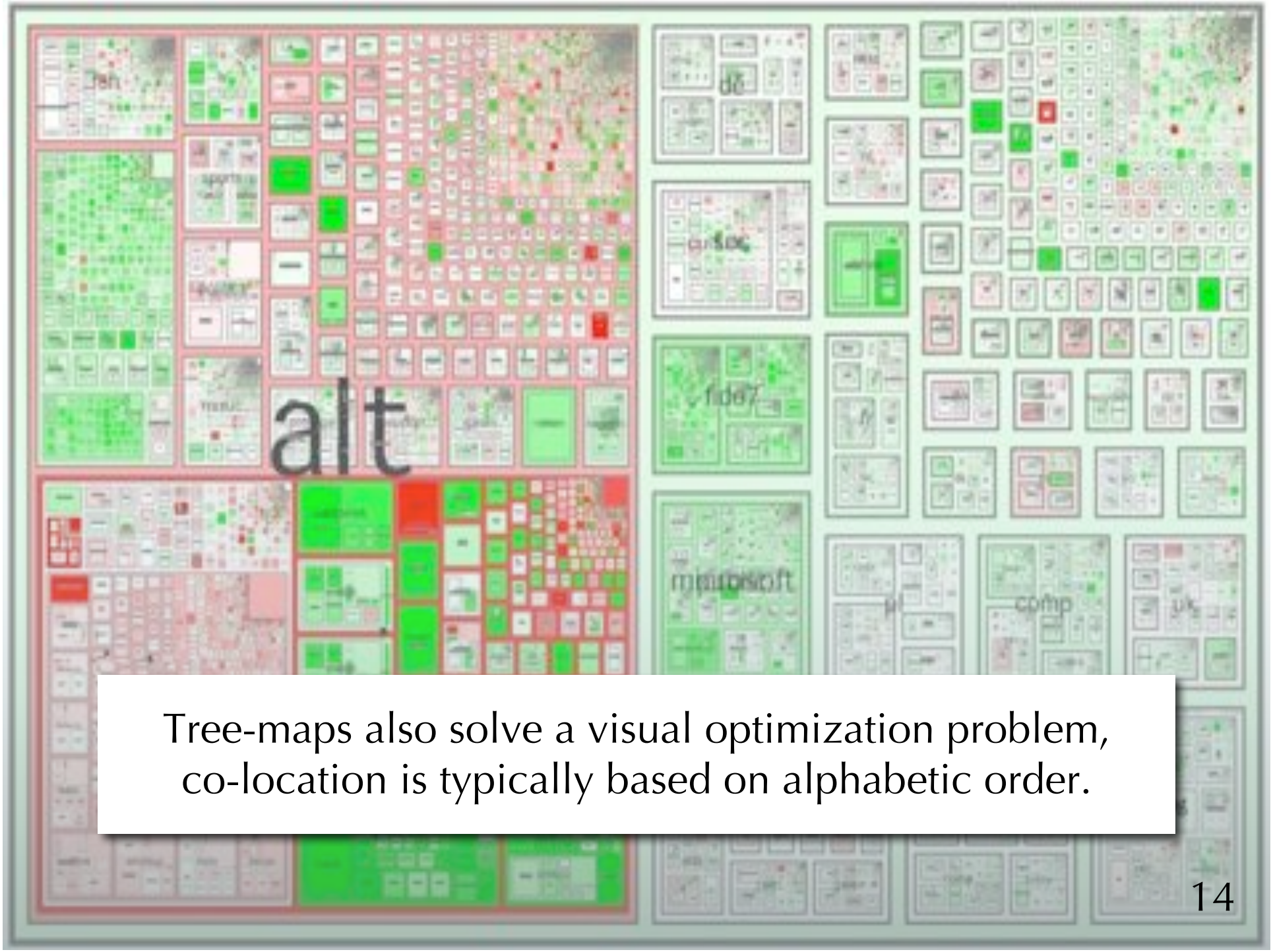


Graph drawing optimizes visual intersections, the co-location of elements is not based on their properties.

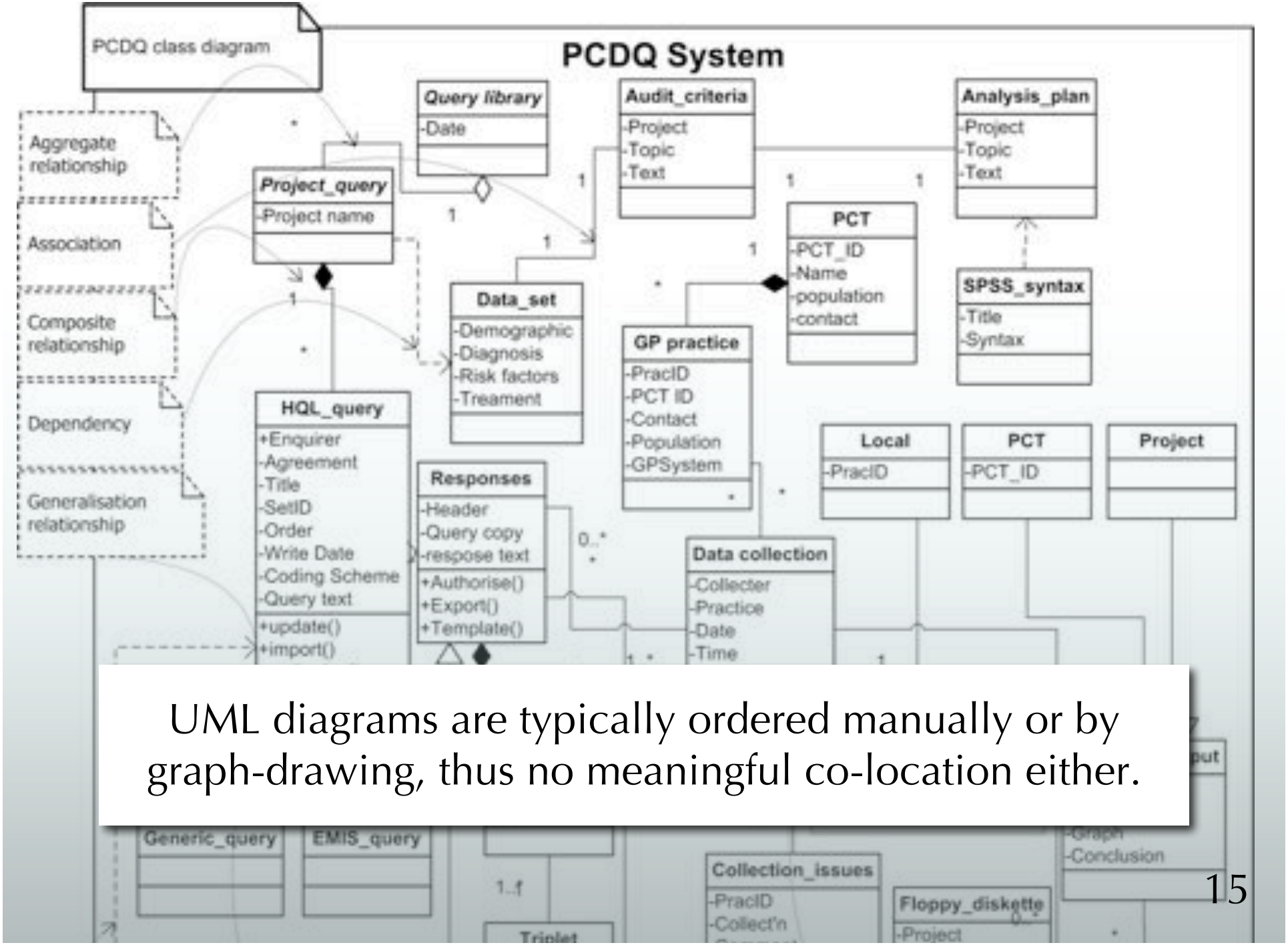


Why are these two near each other, and not lets say "C" and "B"? Not because of common properties but just to avoid crossing!

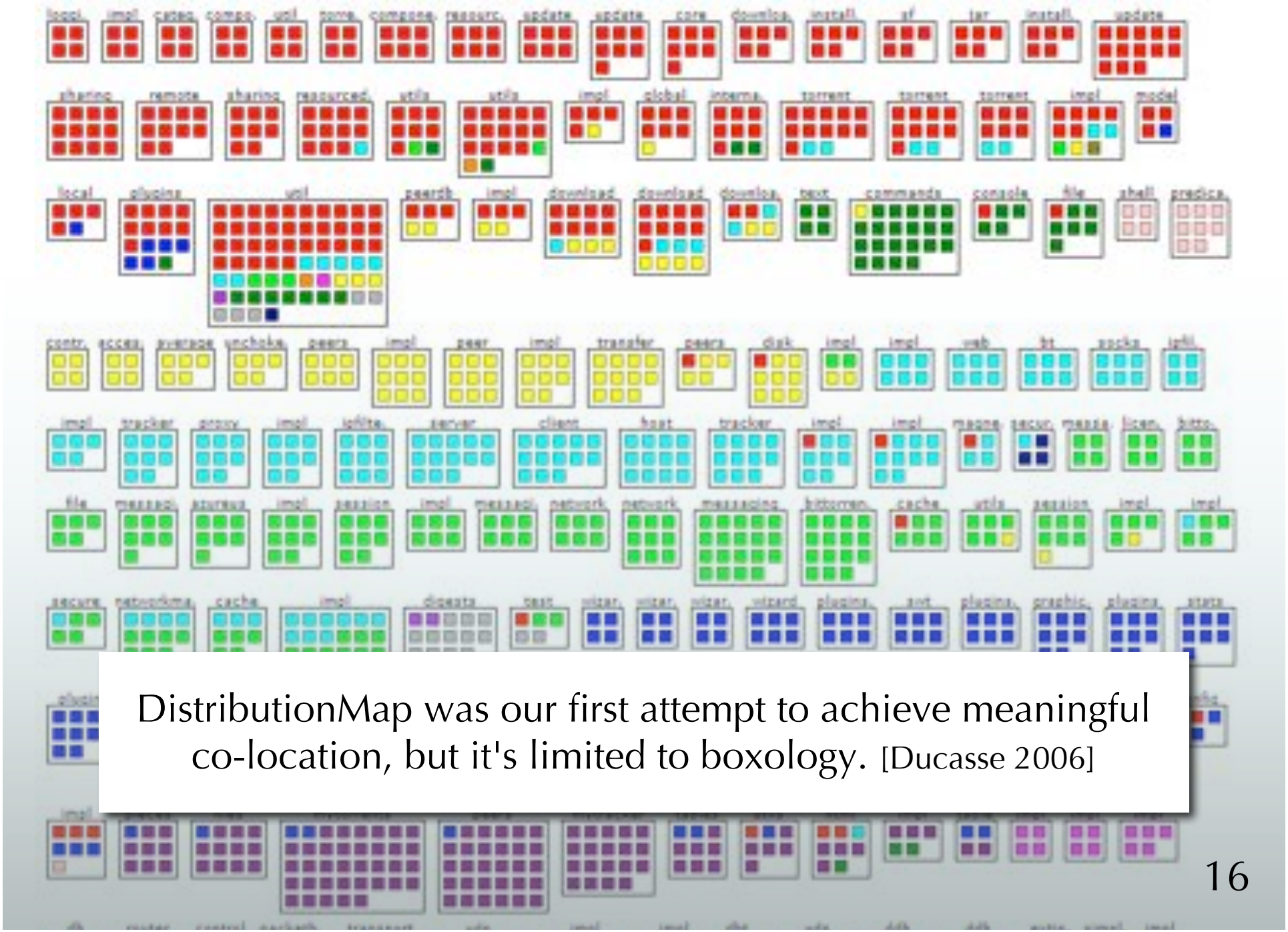
Graph drawing optimizes visual intersections, the co-location of elements is not based on their properties.



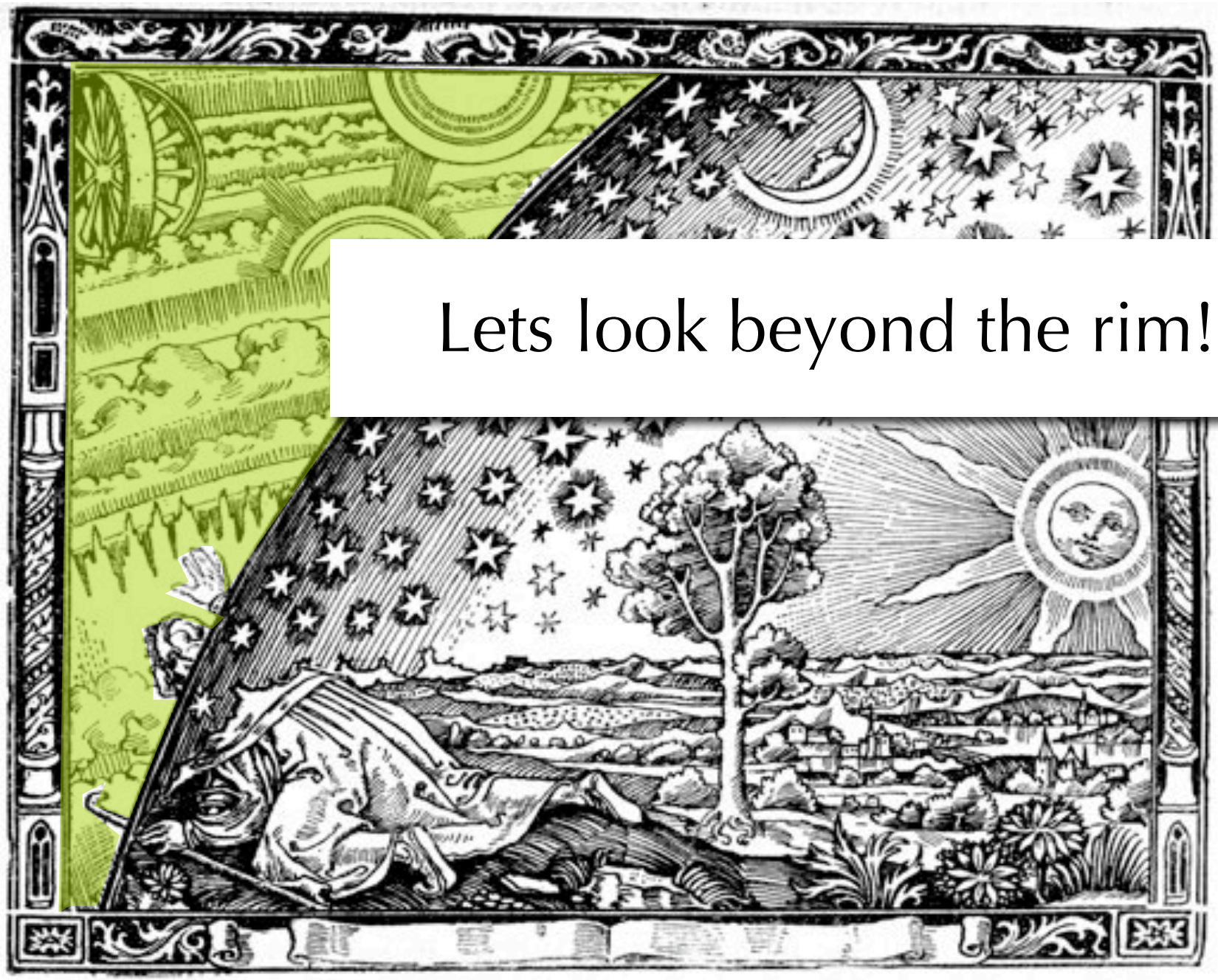
Tree-maps also solve a visual optimization problem, co-location is typically based on alphabetic order.

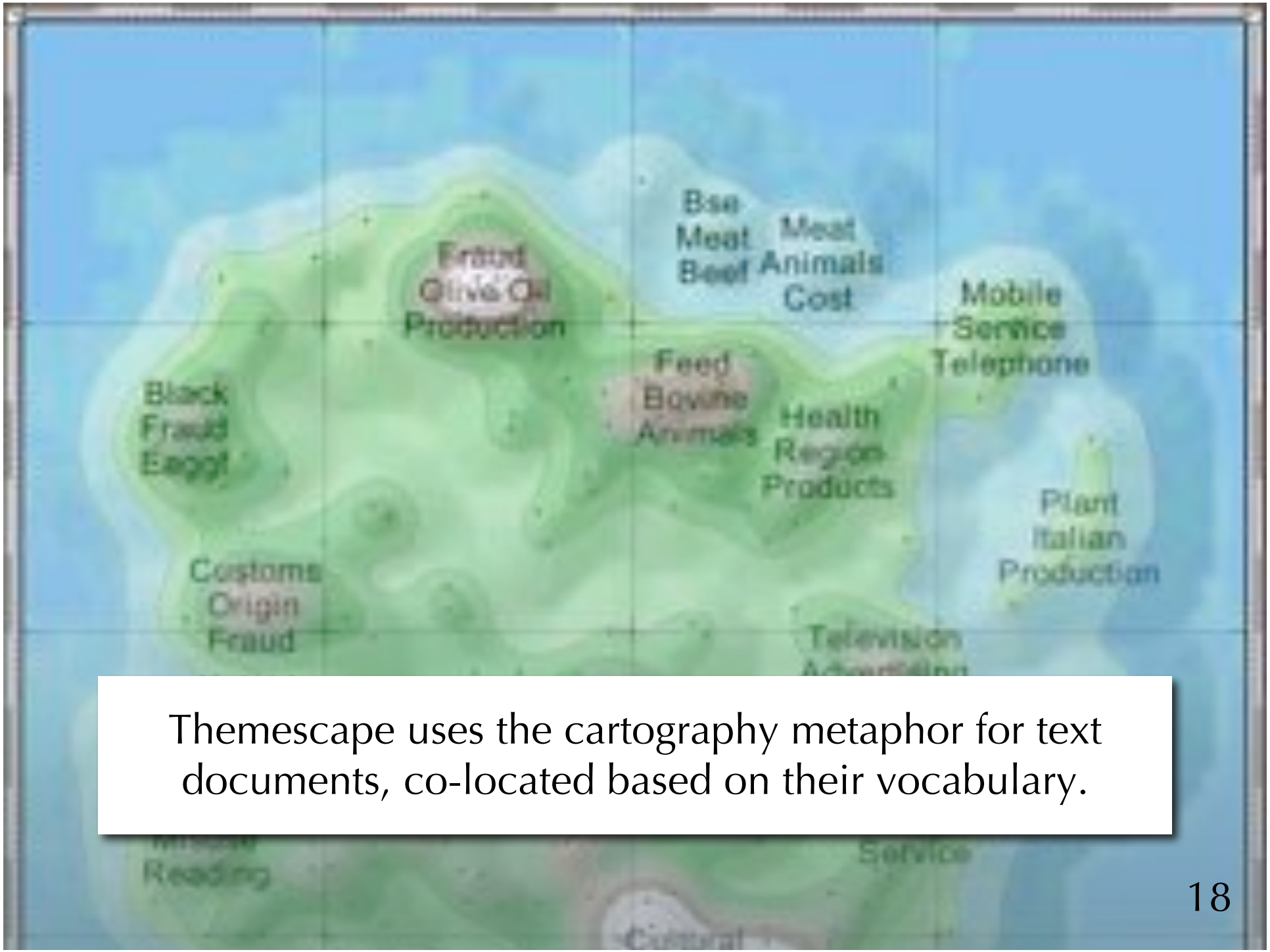


UML diagrams are typically ordered manually or by graph-drawing, thus no meaningful co-location either.



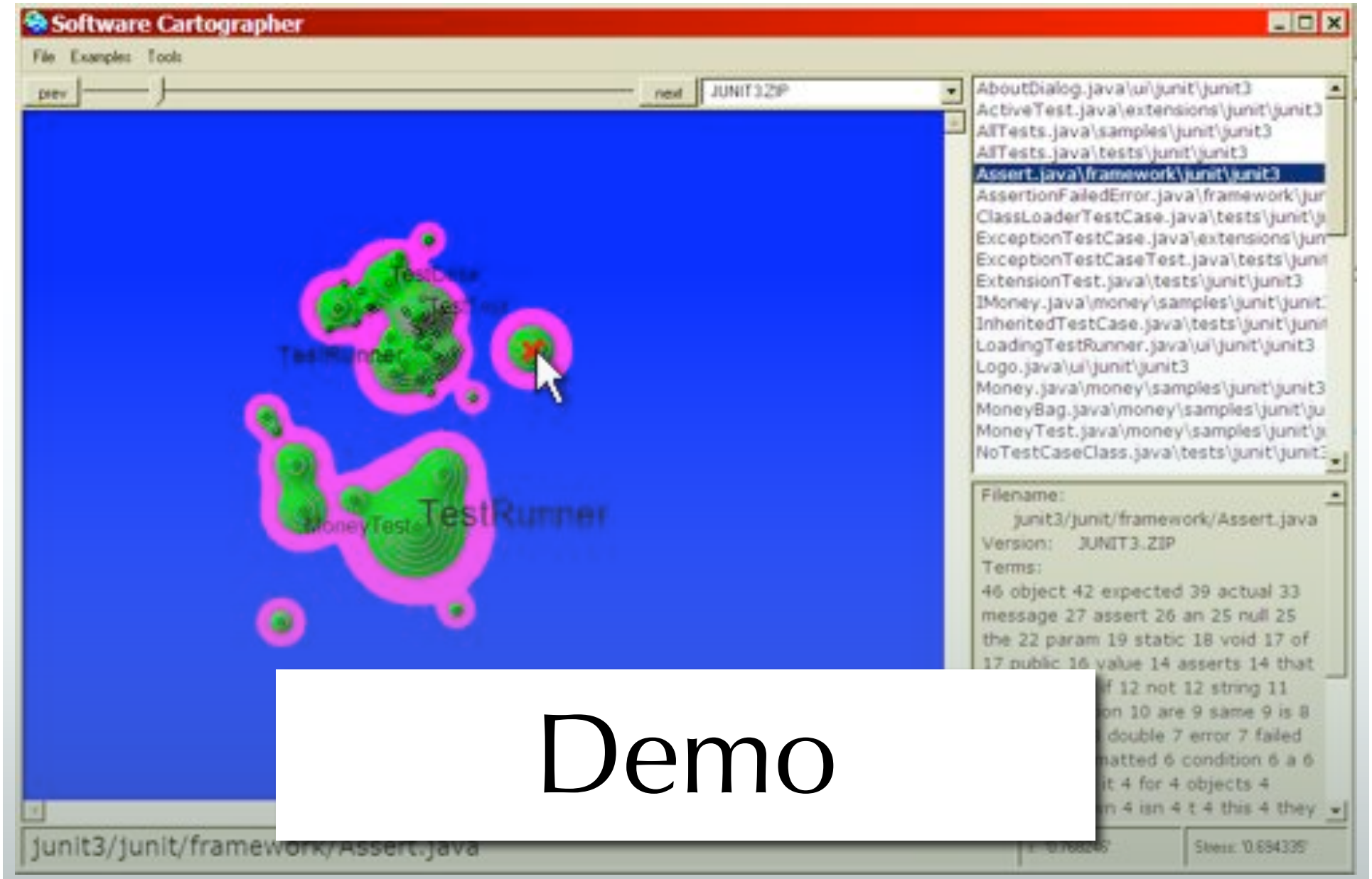
DistributionMap was our first attempt to achieve meaningful co-location, but it's limited to boxology. [Ducasse 2006]


Lets look beyond the rim!



Themescape uses the cartography metaphor for text documents, co-located based on their vocabulary.

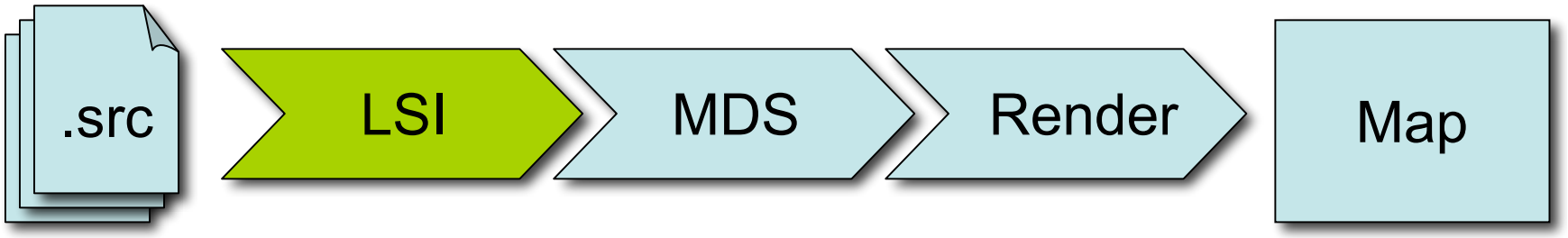
Can this be done for code?


Software Cartographer
File Examples Tools
prev
next
JUNIT3.ZIP
AboutDialog.java\ui\junit\junit3
ActiveTest.java\extensions\junit\junit3
AllTests.java\samples\junit\junit3
AllTests.java\tests\junit\junit3
Assert.java\framework\junit\junit3
AssertionFailedError.java\framework\jur
ClassLoaderTestCase.java\tests\junit\j
ExceptionTestCase.java\extensions\jun
ExceptionTestCaseTest.java\tests\junit
ExtensionTest.java\tests\junit\junit3
IMoney.java\money\samples\junit\junit
InhentedTestCase.java\tests\junit\junit
LoadingTestRunner.java\ui\junit\junit3
Logo.java\ui\junit\junit3
Money.java\money\samples\junit\junit3
MoneyBag.java\money\samples\junit\ju
MoneyTest.java\money\samples\junit\j
NoTestCaseClass.java\tests\junit\junit
TestRunner
MoneyTest
Filename:
junit3/junit/framework/Assert.java
Version: JUNIT3.ZIP
Terms:
46 object 42 expected 39 actual 33
message 27 assert 26 an 25 null 25
the 22 param 19 static 18 void 17 of
junit3/junit/framework/Assert.java

Demo

Why Vocabulary?

- Abstracts away from structure [Maletic 2001]
- Useful for software analysis [Marcus 2004]
- Metric distance between code [Poshyvanyk 2006]
- Clusters software by topic [Kuhn 2007]
- Over time, software grows… [Vasa 2007]
- …but vocabulary remains stable [Antoniol 2007]

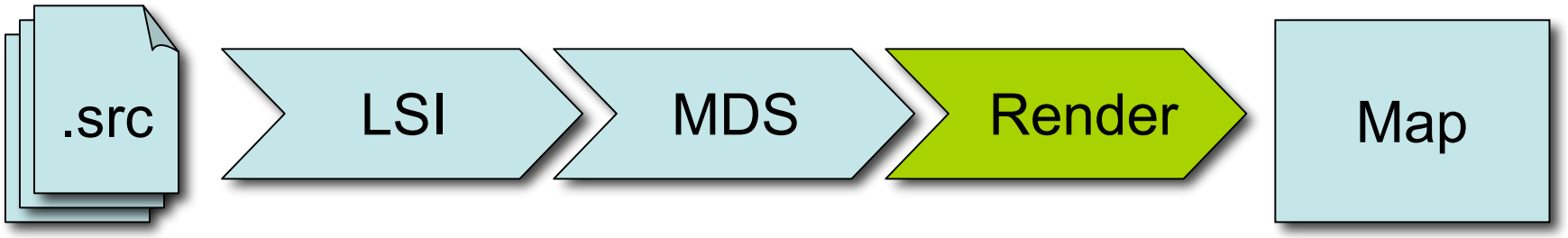


Semantic Clustering [Kuhn 2007]

- Analyses the **vocabulary** of source files.
- Latent Sematic Indexing (LSI) [Deerwester 1990]
- Maps files into n-dim vector space.
- Files with similar vocabulary are close.

Multidimensional Scaling

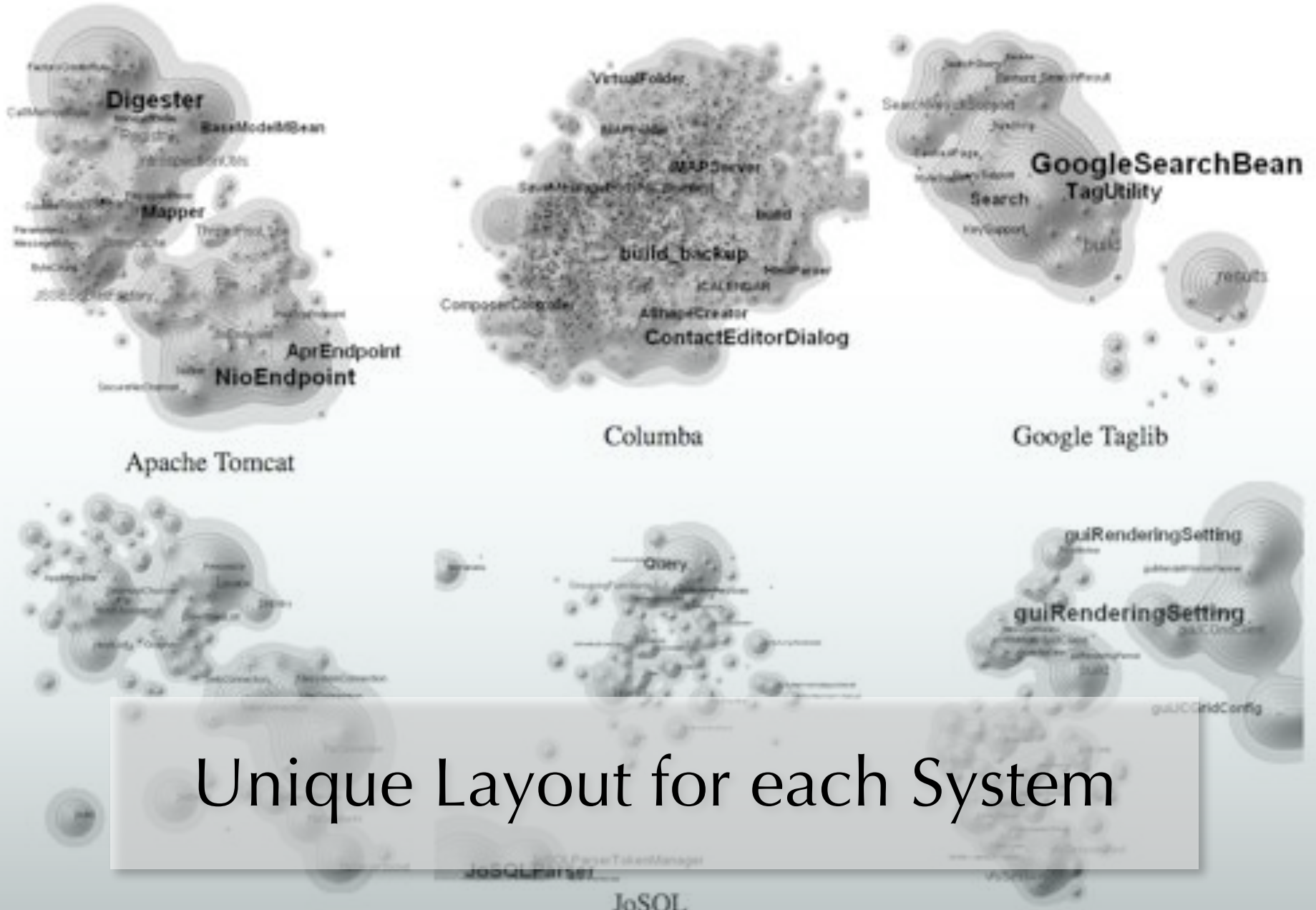
- Maps vectors from n-dim to 2D.
- MDS attempts to preserve the relative distance between vectors.
- Iterative, non-deterministic approximation.



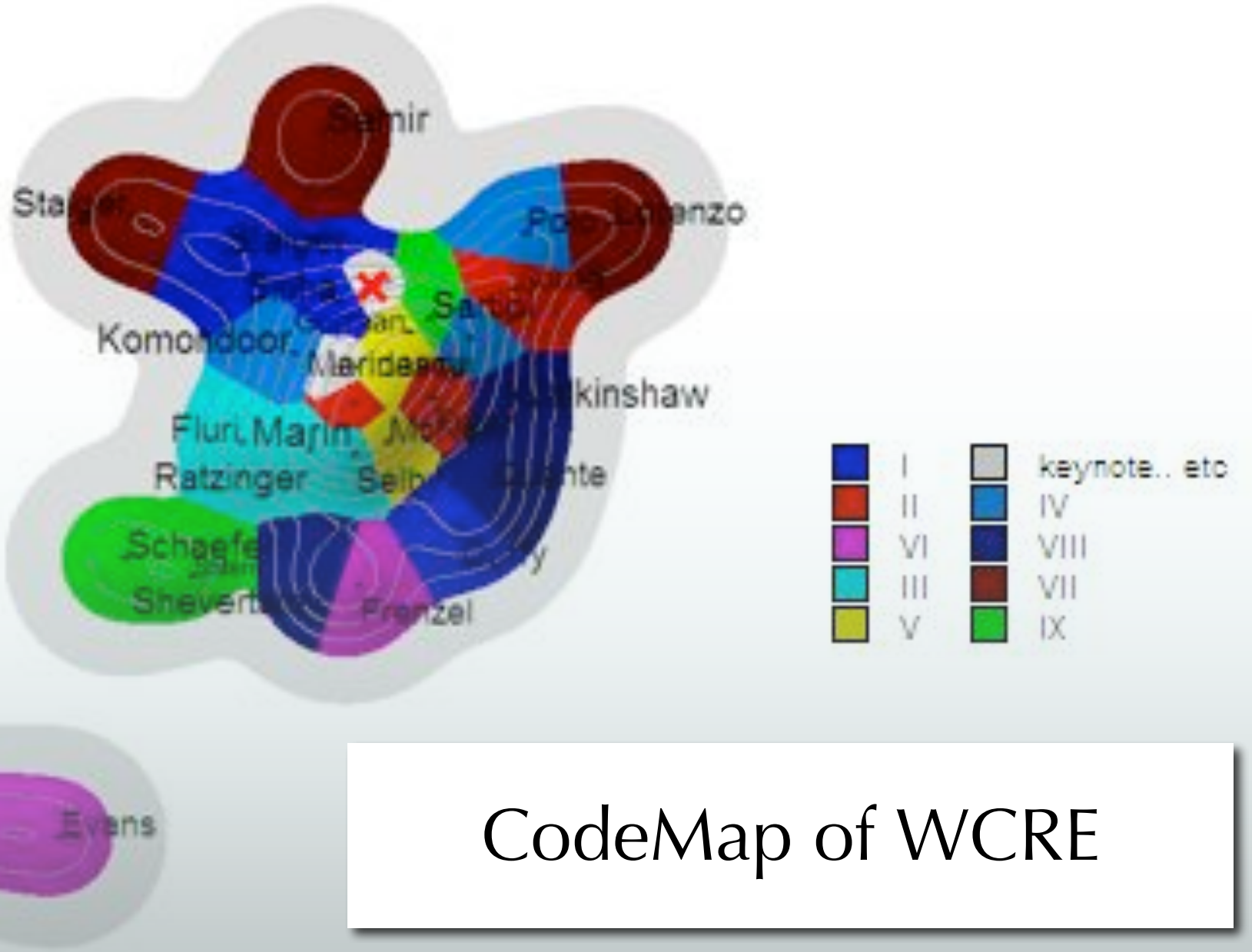
Map Rendering

- Each file is rendered as a hill.
- Hill height equals file size.
- Standard algorithms from cartography:
  - Digital elevation model (DEM)
  - Hill Shading
  - Contour lines

Consistent Layout over Time



Unique Layout for each System



CodeMap of WCRE