

Η Εξόρυξη Διαδικασιών ως Εργαλείο Αποφάσεων



Αναπληρωτής Καθηγητής Παύλος Δελιάς,
Τμήμα Λογιστικής & Χρηματοοικονομικής
pdelias@teiemt.gr, @PavlosDelias





Fontaine Maury

									and Nimbus.			
Dec.	1	Midn't.	31 52	50 51			N. N. W.	Light breeze.	Nimbus and Cum. Stratus.	10	S. W.	30.10
		8 A. M.	31 59	49 48			S. W.	Moderate.	Nimbus and Cirro Cum. Stratus.	10	S'd. & W'd.	30.12
		4 P. M.	32 00	49.13			S. S. E.	Light breezes.	Nimbus.	10	Southward.	30.12
"	2	Midn't.	32 00	48 50			S. S. E.	Light breezes.	Nimbus.	10	Southward.	30.12
		8 A. M.	32 24	48 26	N. 53° W.	0.5	S. E.	Moderate.	Cum. Stratus.	8	S'd. & E'd.	30.22
		4 P. M.	32 25 20	47 50	Westward.	0.5	South.	Moderate.	Cir. Cum. Strat.	5	Southward.	30.18
"	3	Midn't.	32 25	47 1			S. by E.	Moderate.	Cum. Stratus.	6	Southward.	30.24

6	5	"	Small rain, wind	
7	7	"		
8	7	"		11
9	8	"	MC. Pearl Main Bay	
10	8	"		
11	8	"	incessant rain, with hard	
12	8	"	gale.	P.S.
			178 Miles.	

No observations.

Course & Dist.	Lat	Long	Lat	Long	Lat	Long	Lat	Long	Lat	Long
170 E 171	acc 161	acc 161	acc 161	acc 161	acc 161	acc 161	acc 161	acc 161	acc 161	acc 161

Depart. 60 E of Acc. v. of Acc. v. Bar. Ther. 68°

1	8	4	SE breeze, fresh	
2	9	"	heavy rain, & squalls.	
3	9	"	Same as yesterday.	
4	9	"		6
5	9	"	North breeze, all day.	
6	9	4	light rain, & squalls.	
7	10	4	Strong breeze with hard	
8	10	4	gale & incessant rain.	6
9	10	"	Heavy rain.	
10	10	4	Squally	
11	10	4	Hard rain.	
12	10	"	Heavy rain.	6 1/2
1	10	"		
2	10	4	North breeze, all day.	
3	10	4		
4	10	"	SE	6 1/2
5	9	4		
6	10	"		
7	9	4	Fine?	
8	13	"	Outrigger & made all sail.	7
9	7	4		
10	7	4		

Sunday 18th April 1830.

First had fresh breeze @ NE. drawing to the North & NW in the middle, with incessant heavy rain, & hard squalls: latterly moderate breeze @ NW & cloudy weather.

All performed Divine Service in the Sticage, the unsettled weather preventing it on Deck. At Noon mustered the Ship's Comp^y.

Sick List 9 Seamen.

Water expended this week Gallons 1061

EXPLANATION

-  North-East Trade
-  South-East Trade
-  South-East & North-West Monsoon
-  North-East & South-West Monsoon
-  Favouring Westerly Winds

Άνεμοι

Ιδανικές Διαδρομές

-  Arrow a vessel passing to starboard
-  Fish a vessel
-  Boat a vessel

-  Line of distance for
-  Arrow route
-  Squat



©PHOTO Rienk Mebius 2007

Δεδομένα Υπάρχουν...

- ERP
- CRM
- BPMS / WfMS
- RFID, GPS
- ...
- "The world's data will grow by 50 times in the next 10 years" (IDC),
- "90% of the data in the world today has been created in the last two years alone" (IBM).

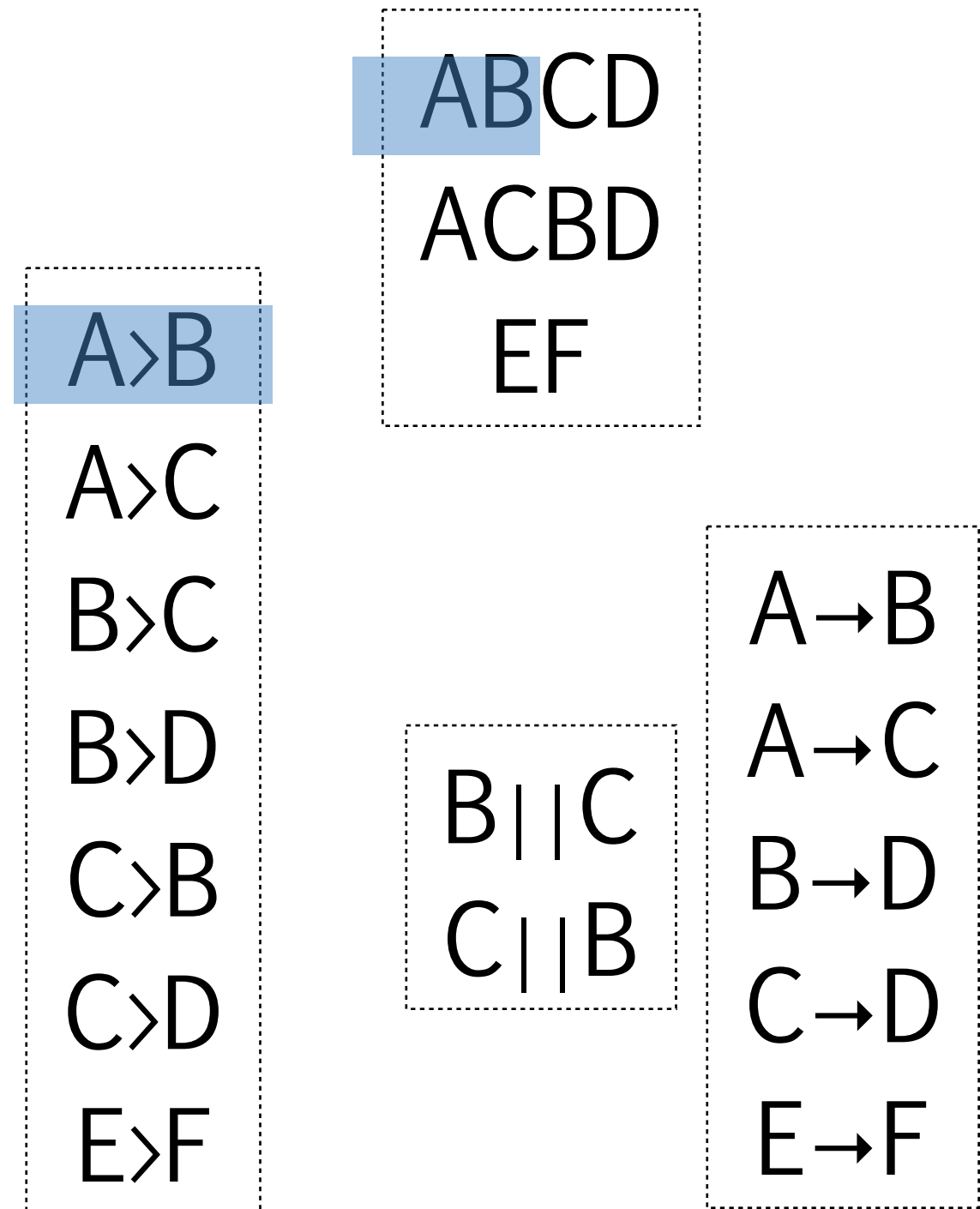
Συλλογή Δεδομένων

	A	B	C
1	CaseID	Event	TimeStamp
2	1	Arrival	1/6/12 21:50
3	1	Assortment	1/6/12 21:55
4	1	Diagnosis	1/6/12 21:57
5	1	Blood_Test	1/6/12 22:05
6	1	Blood_Test	1/6/12 22:45
7	1	Biochem_Test	1/6/12 22:05
8	1	Biochem_Test	1/6/12 23:20
9	1	Entrance_to_ER_Room	1/6/12 21:55
10	1	Exit_from_ER_Room	1/6/12 23:35
11	1	ER_Exit	1/6/12 23:35
12	2	Arrival	1/6/12 22:30
13	2	Assortment	1/6/12 22:35
14	2	Diagnosis	1/6/12 22:36
15	2	Blood_Test	1/6/12 22:40
16	2	Blood_Test	1/6/12 23:20
17	2	Biochem_Test	1/6/12 22:40
18	2	Biochem_Test	1/6/12 23:55
19	2	Additional_Test	1/6/12 22:50
20	2	Additional_Test	1/6/12 23:50
21	2	Entrance_to_ER_Room	1/6/12 22:35
22	2	Exit_from_ER_Room	2/6/12 0:09
23	2	Prescription	2/6/12 0:05
24	2	ER_Exit	2/6/12 0:10
25	3	Arrival	1/6/12 23:05
26	3	Assortment	1/6/12 23:10
27	3	Diagnosis	1/6/12 23:12
28	3	Blood_Test	1/6/12 23:15
29	3	Blood_Test	1/6/12 23:55
30	3	Biochem_Test	1/6/12 23:15
31	3	Biochem_Test	2/6/12 0:35
32	3	Additional_Test	1/6/12 23:25
33	3	Additional_Test	2/6/12 0:20
34	3	Entrance_to_ER_Room	1/6/12 23:10
35	3	Exit_from_ER_Room	2/6/12 0:47

- Ένα αρχείο καταγραφών
- Απαραίτητα στοιχεία:
 - Case ID, Event, Timestamp
- Προαιρετικά:
 - Δράστης, Χαρακτηριστικά

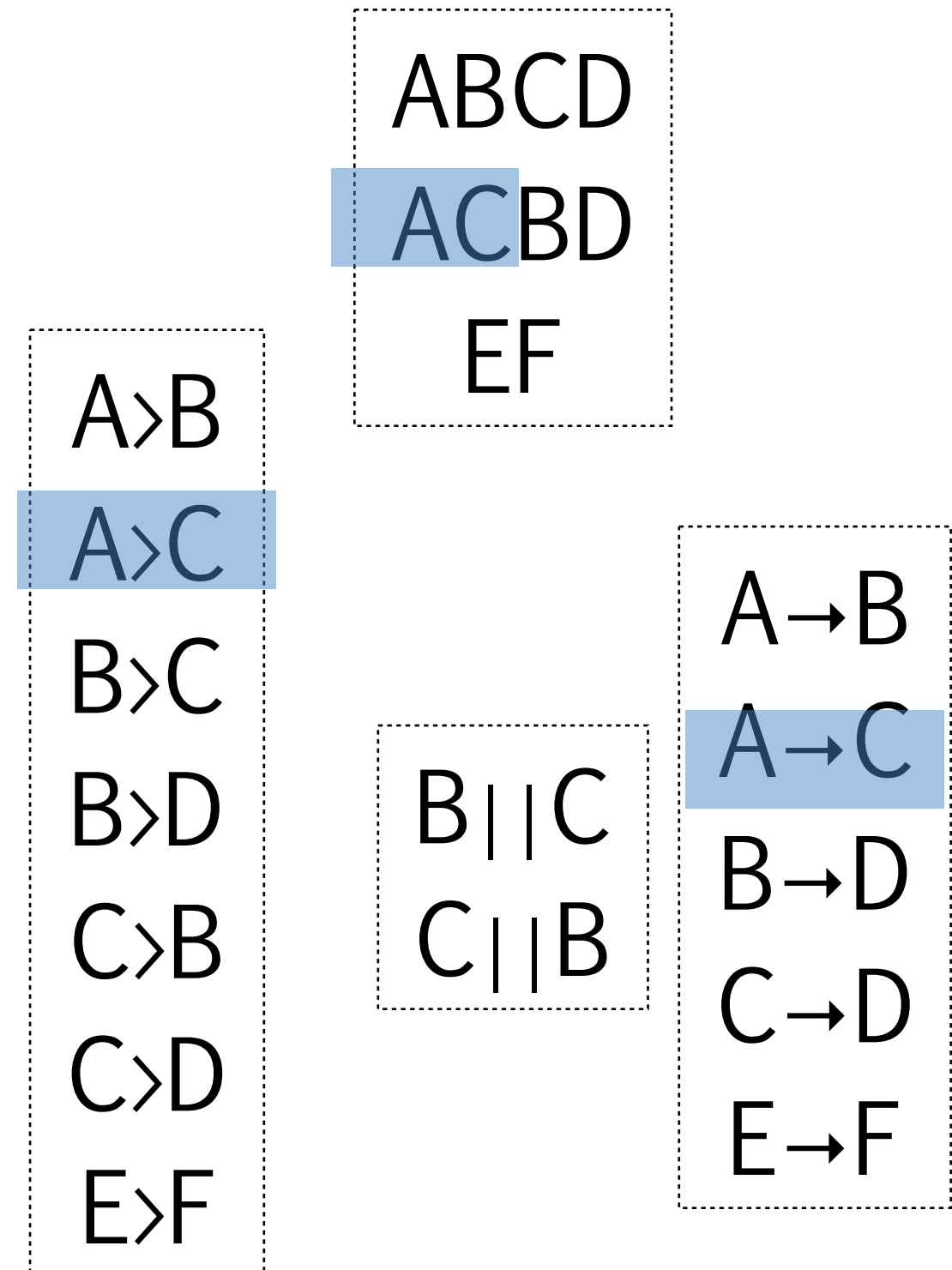
Βασικές έννοιες για την Ανακάλυψη: Ο αλγόριθμος a

- Σχέσεις Διάταξης $>$, $\rightarrow$, $\parallel$, $\#$
- **Άμεση Διαδοχή:** $x > y$ αν και μόνο αν για κάποια περίπτωση το x ακολουθείται άμεσα από το y
- **Αιτιότητα:** $x \rightarrow y$ αν και μόνο αν $x > y$ και όχι $y > x$
- **Παραλληλία:** $x \parallel y$ αν και μόνο αν $x > y$ και $y > x$
- **Μη σχετισμένα:** $x \# y$ αν και μόνο αν όχι $x > y$ και όχι $y > x$



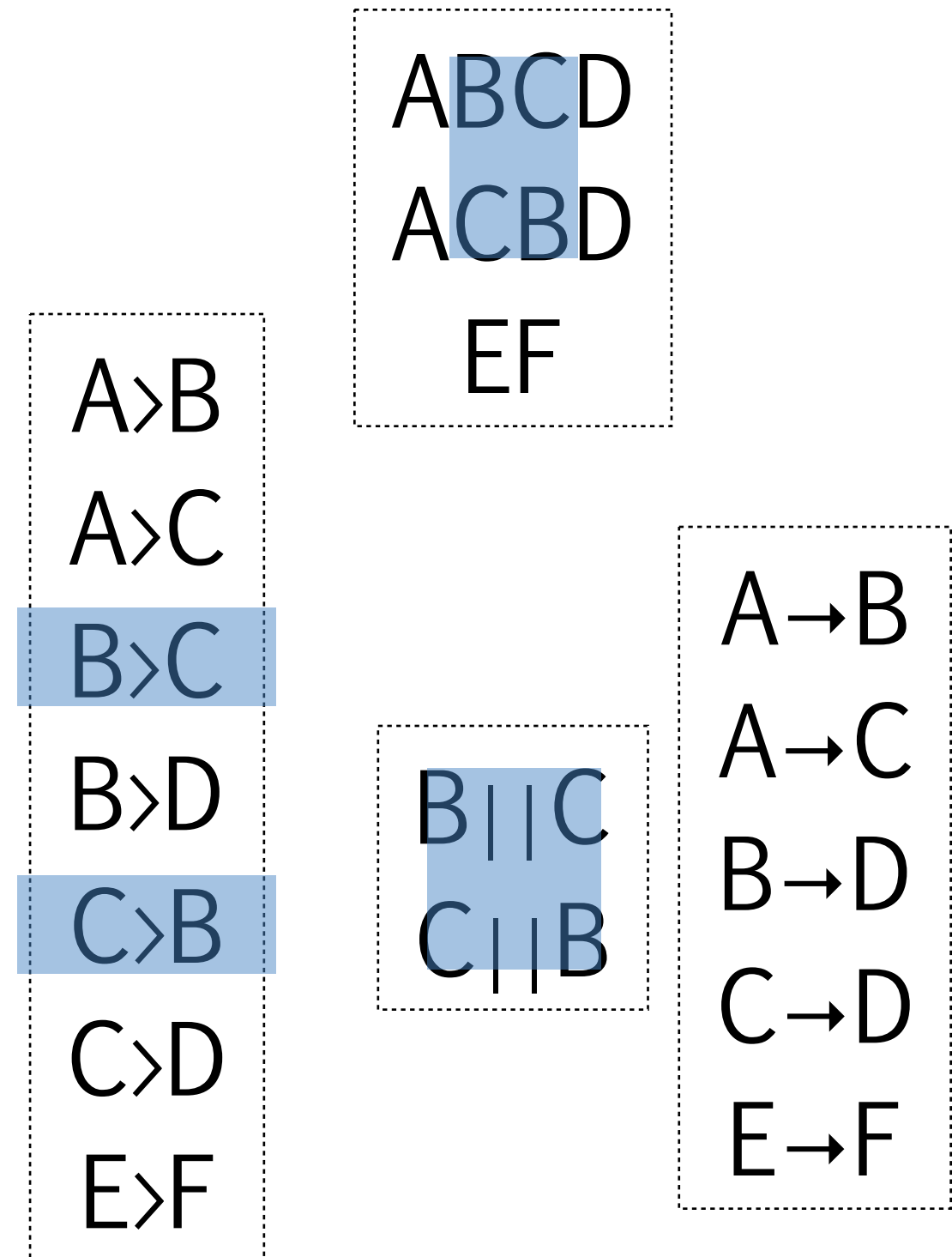
Βασικές έννοιες για την Ανακάλυψη: Ο αλγόριθμος a

- Σχέσεις Διάταξης $>$, $\rightarrow$, $\parallel$, $\#$
- **Άμεση Διαδοχή:** $x > y$ αν και μόνο αν για κάποια περίπτωση το x ακολουθείται άμεσα από το y
- **Αιτιότητα:** $x \rightarrow y$ αν και μόνο αν $x > y$ και όχι $y > x$
- **Παραλληλία:** $x \parallel y$ αν και μόνο αν $x > y$ και $y > x$
- **Μη σχετισμένα:** $x \# y$ αν και μόνο αν όχι $x > y$ και όχι $y > x$

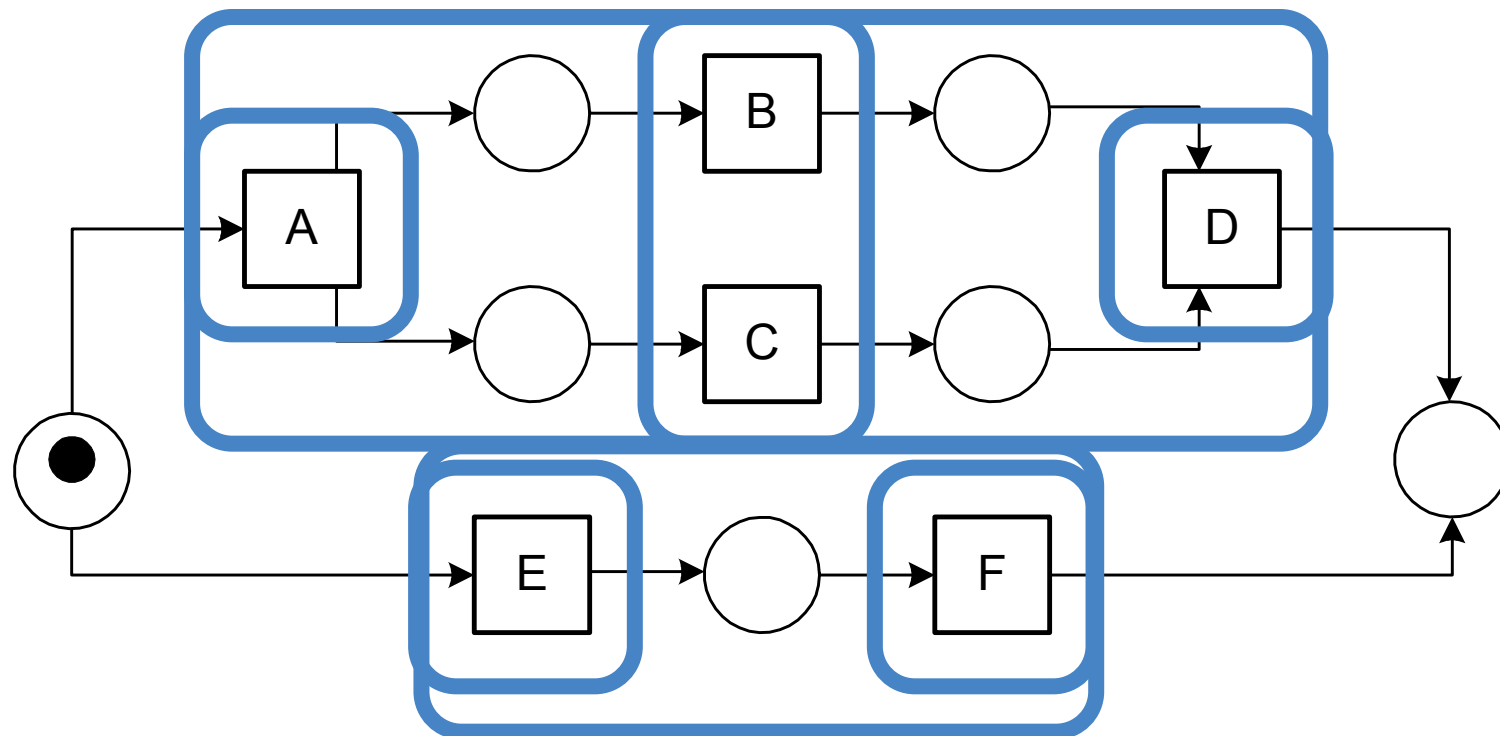
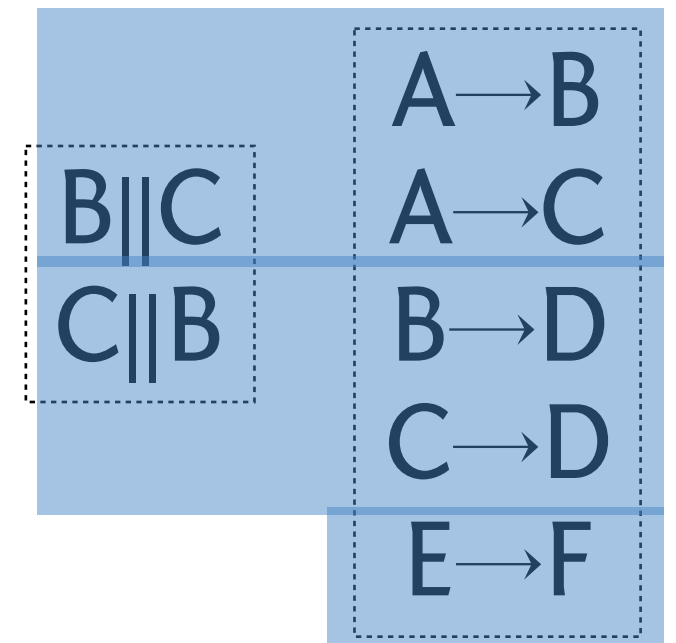
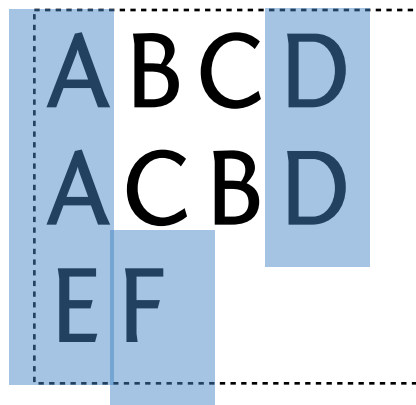


Βασικές έννοιες για την Ανακάλυψη: Ο αλγόριθμος a

- Σχέσεις Διάταξης $>$, $\rightarrow$, $\parallel$, $\#$
- **Άμεση Διαδοχή:** $x > y$ αν και μόνο αν για κάποια περίπτωση το x ακολουθείται άμεσα από το y
- **Αιτιότητα:** $x \rightarrow y$ αν και μόνο αν $x > y$ και όχι $y > x$
- **Παραλληλία:** $x \parallel y$ αν και μόνο αν $x > y$ και $y > x$
- **Μη σχετισμένα:** $x \# y$ αν και μόνο αν όχι $x > y$ και όχι $y > x$

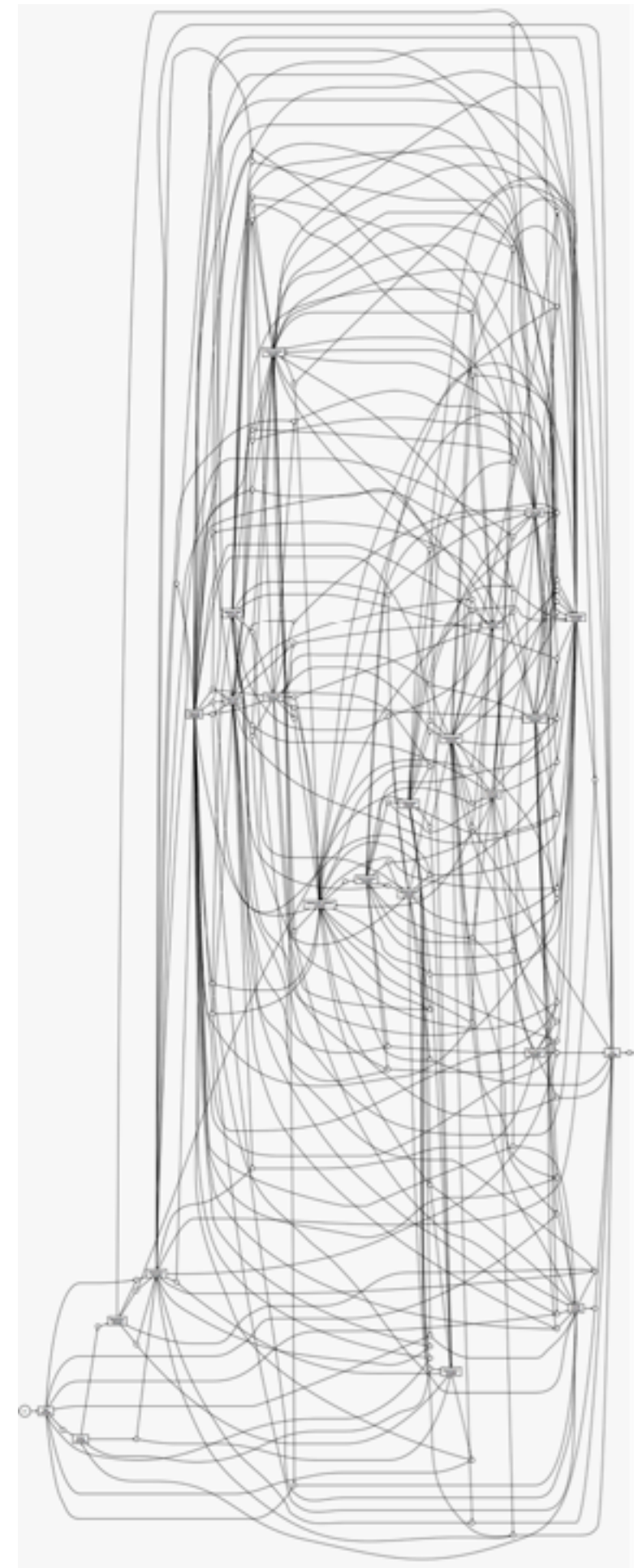


Ο αλγόριθμος a (συνέχεια...)



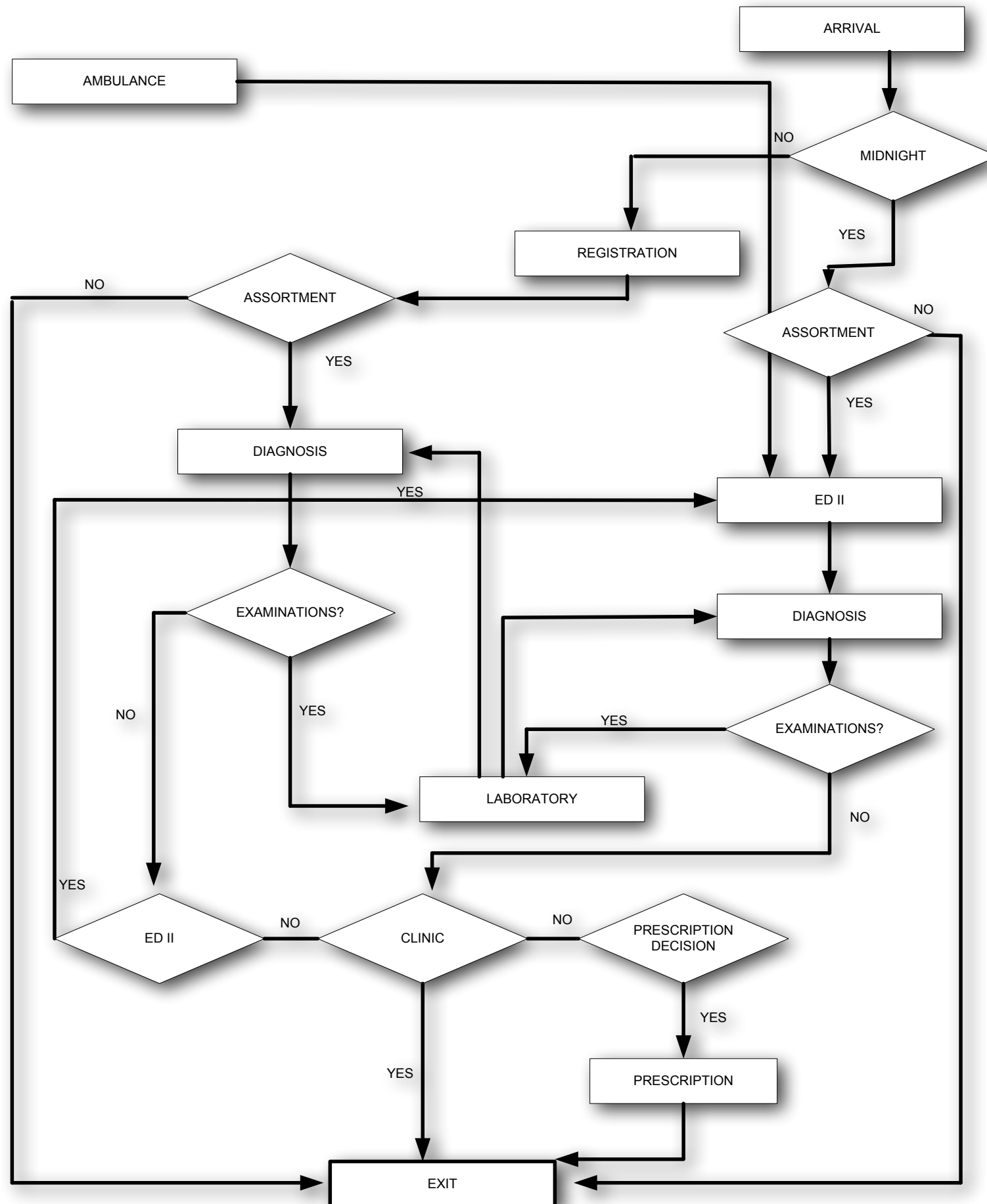
Αλλά...

- Spaghetti (αχρειαστα πολύπλοκο)
- Περιορισμοί σχετικά με:
 - Σύντομους κύκλους
 - Μή-τοπικές εξαρτήσεις
- Αστάθεια
- Μεγάλος υπολογιστικός φόρτος

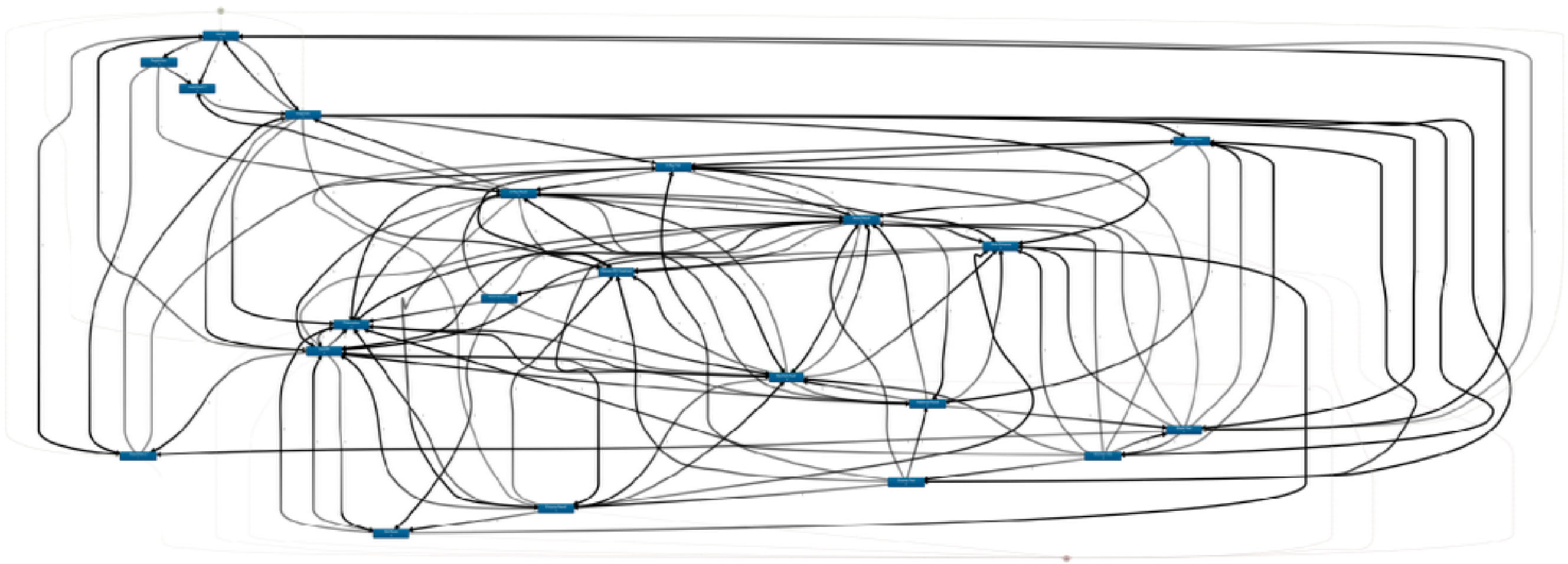


Προκλήσεις

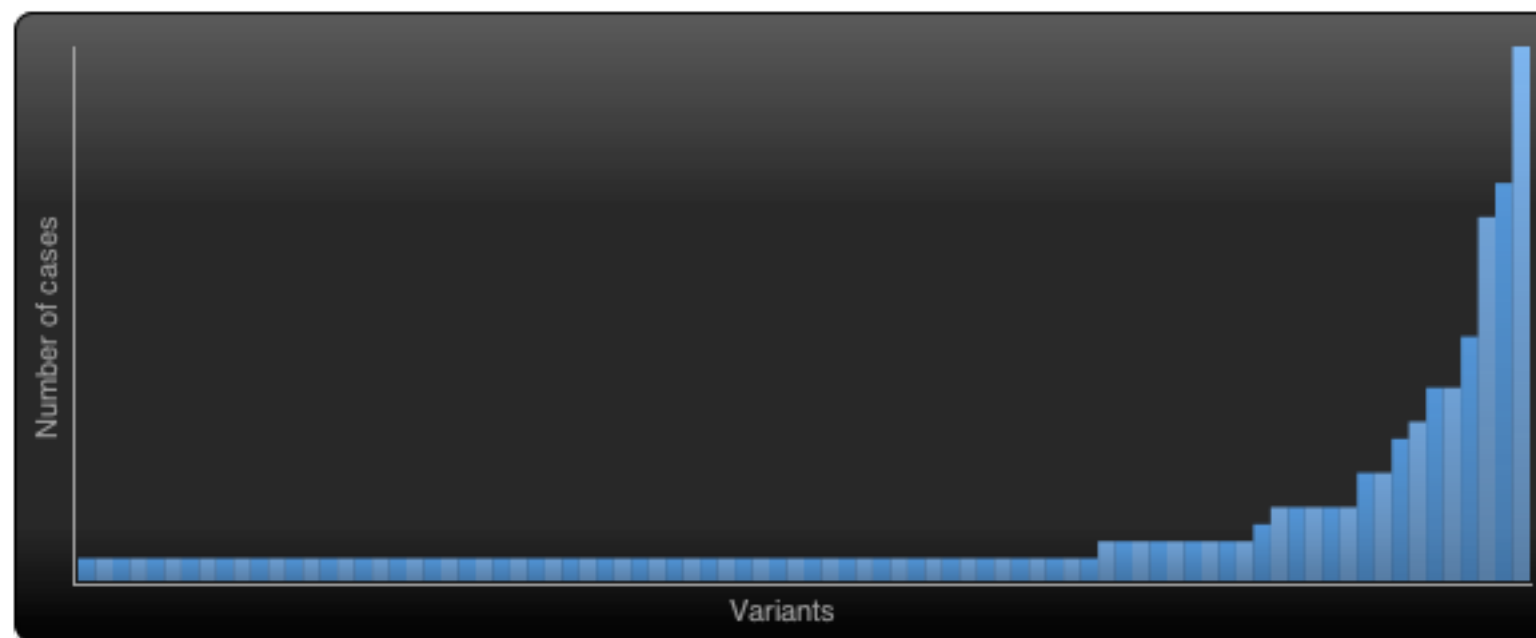
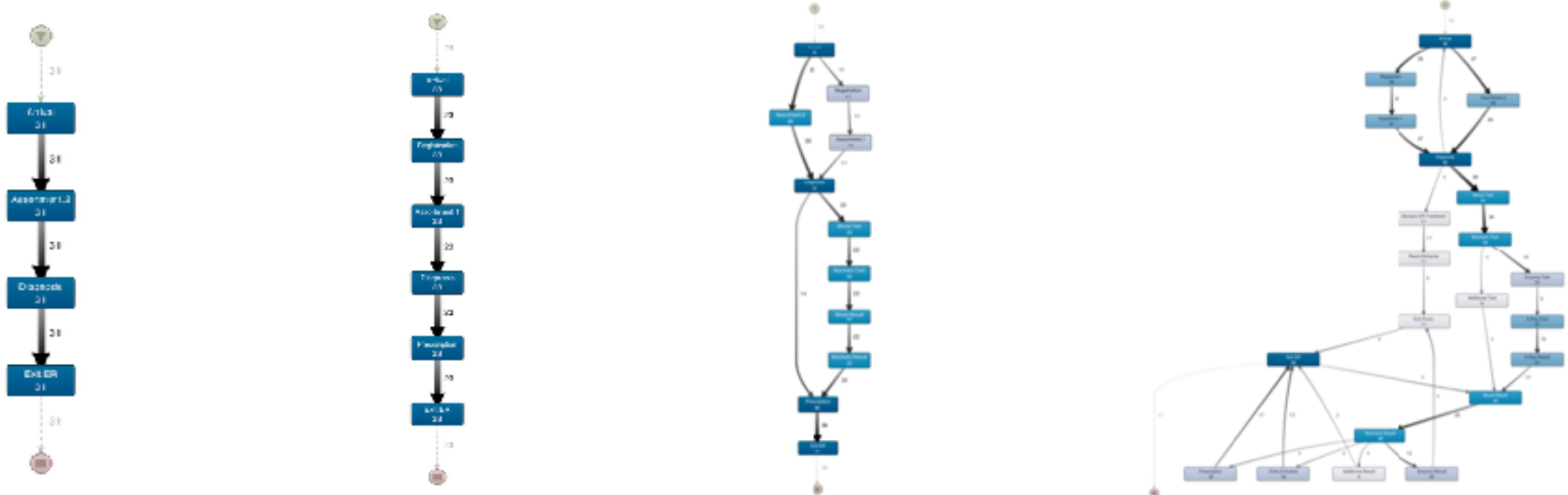
- Ο αριθμός των περιπτώσεων στο αρχείο καταγραφών
- Ο μέσος αριθμός γεγονότων ανά περίπτωση
- Ο αριθμός των διακριτών γεγονότων
- Ο αριθμός των διακριτών διαδρομών



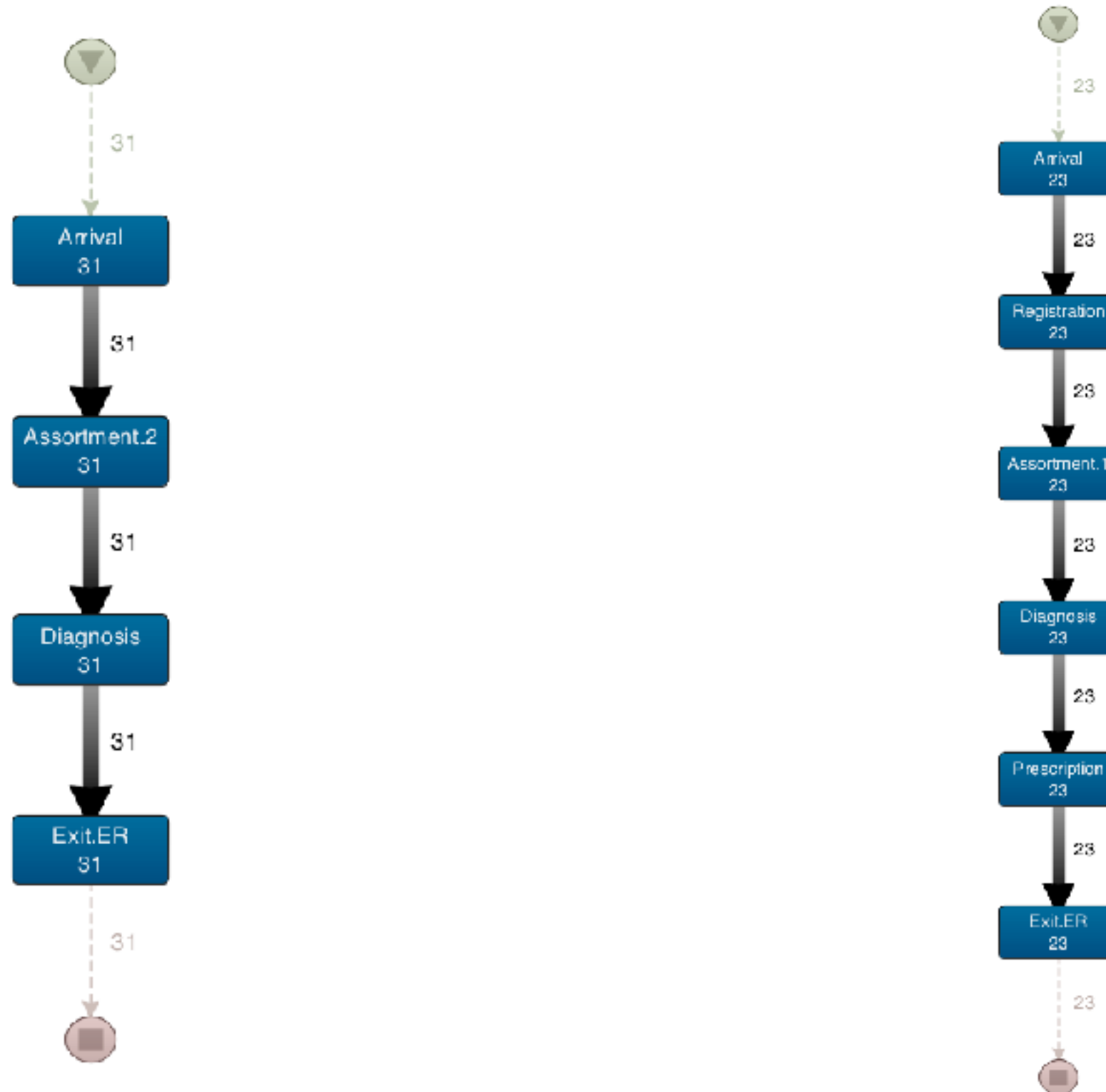
Ο αριθμός των διακριτών διαδρομών (παραλλαγές)



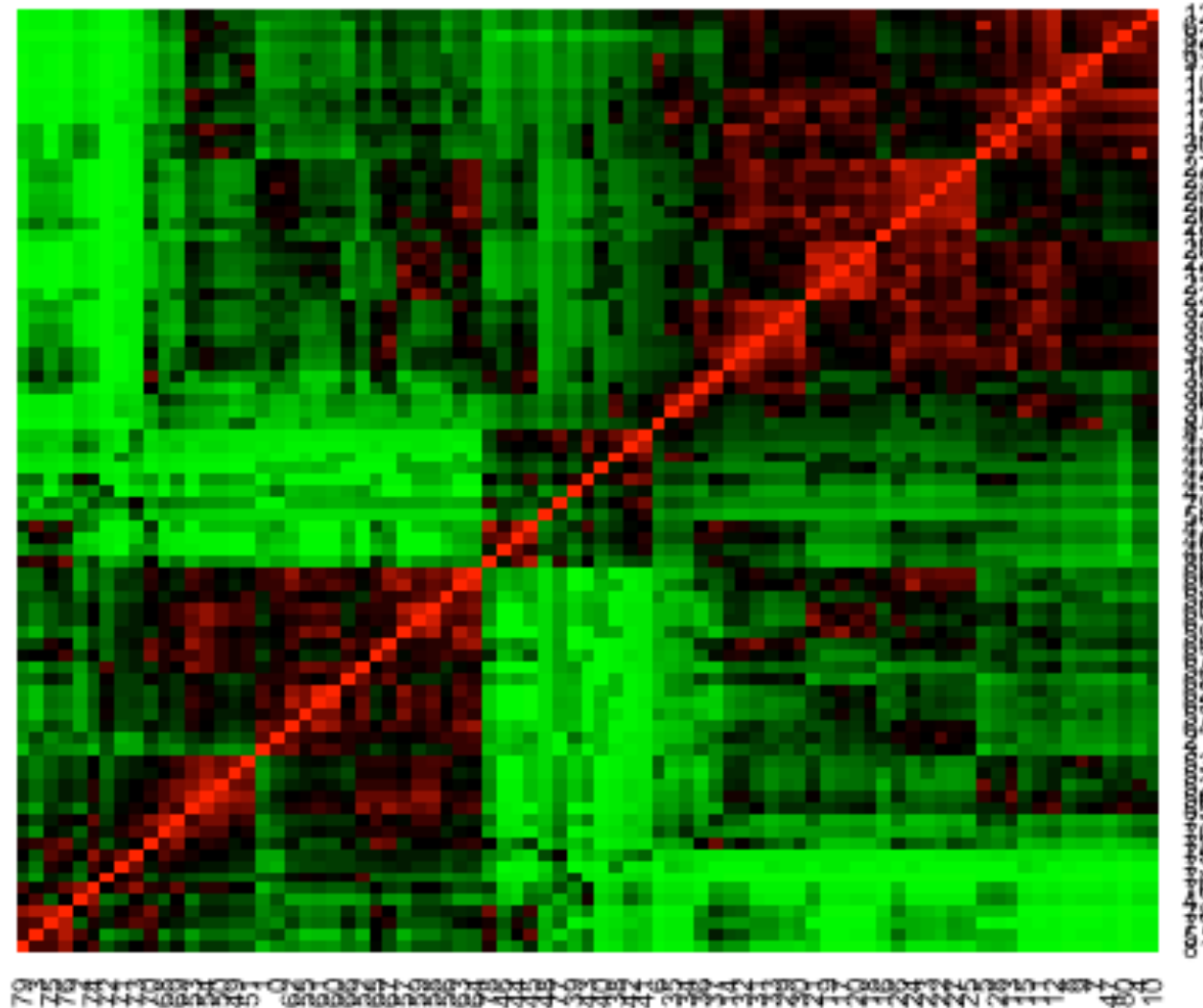
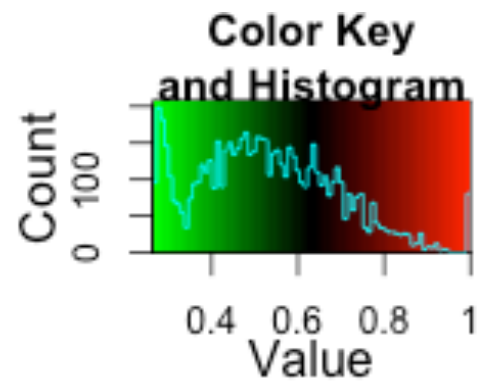
Υψηλή Μεταβλητότητα Ροών



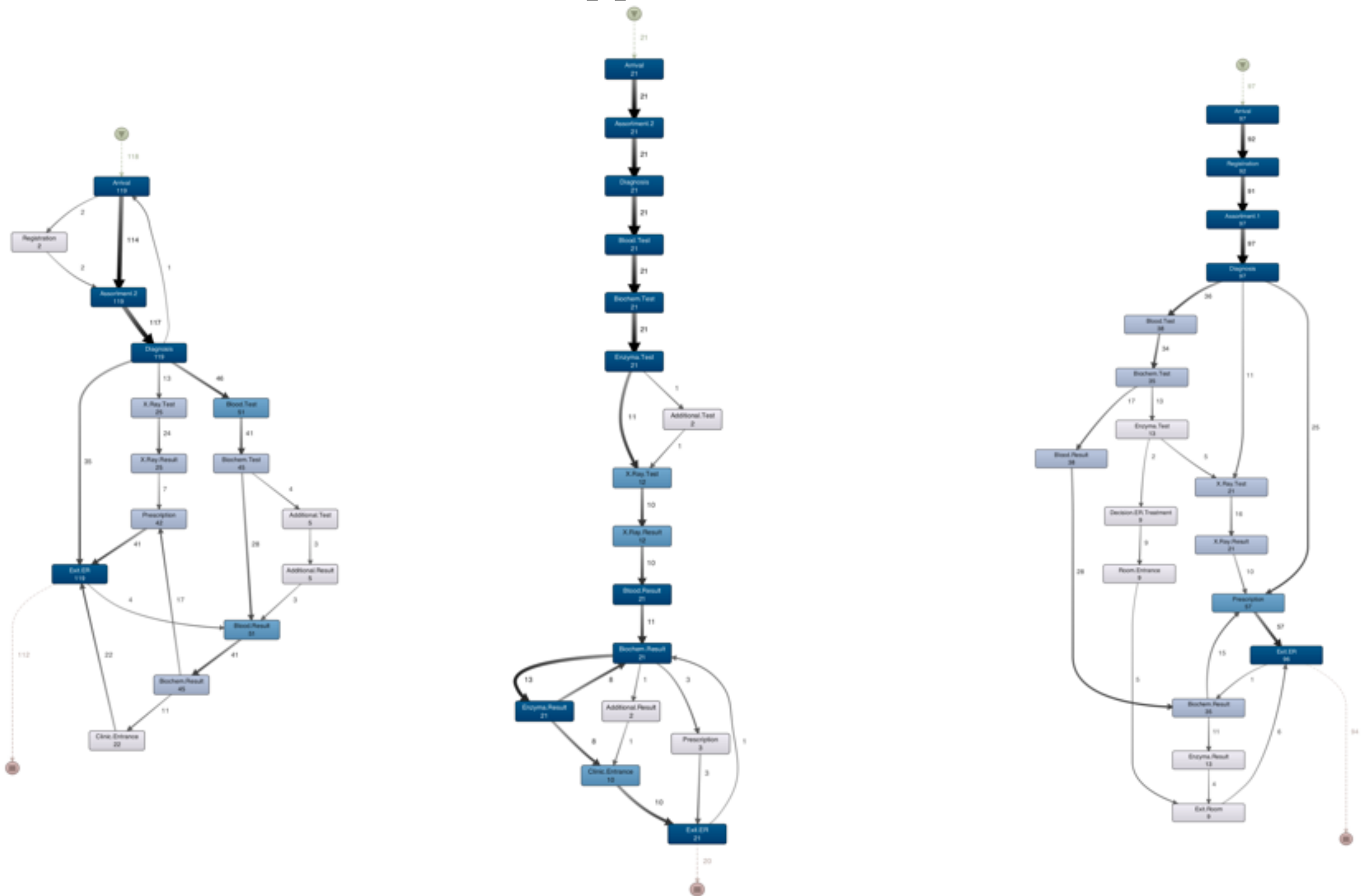
Υπάρχουν ομοιότητες;



Πίνακας Ομοιοτήτων



Τώρα, βγαίνει κάποιο νόημα...



Γιατί είναι σημαντικό; (μελέτη περίπτωση νοσοκομείου)

- Εκ των υστέρων, μπορέσαμε να συνδέσουμε τις ομάδες των ασθενών με τη σοβαρότητα των περιστατικών (triage)
- Επικοινωνία των λειτουργικών παραμέτρων στους ιατρούς.
- Χρήσιμα συμπεράσματα (π.χ. οικονομική κρίση και νυχτερινοί επισκέπτες)

Ανάλυση Πολιτικών Γεγονότων με Εξόρυξη Διαδικασιών

How events are registered?

Reports claim Trump urged Comey to end Michael Flynn investigation

A congressional committee has requested all FBI notes of meetings between Trump and Comey



7:35 17 May 2017
Stephen McNeice
07:35 Wednesday 17 May 2017

Donald Trump is facing fresh controversy after US media reported that he urged the then FBI director James Comey to end an investigation into former national security adviser Michael Flynn.



The story was **first reported in the New York Times**, in a story detailing an apparent memo written by Mr Comey after a meeting in the Oval Office in February.

The story was later reported by other US media outlets, **including CNN**.

According to the reports, Mr Trump told Mr Comey: "I hope you can let this go."

Michael Flynn **resigned from his position in February**, amid claims he misled the administration about his contacts with Russia.

The FBI, meanwhile, has been carrying out an investigation into alleged links between the Trump campaign and Russia.

SQLDATE	20170510
SOURCE ACTOR CODE	USAGOV
TARGET ACTOR CODE	RUS
EVENT CODE	064 (Share intelligence or information)
Avg. Tone	-2.118

Event Coding Frameworks

- WEIS
- COPDAB
- ICEWS
- **CAMEO**
- ...

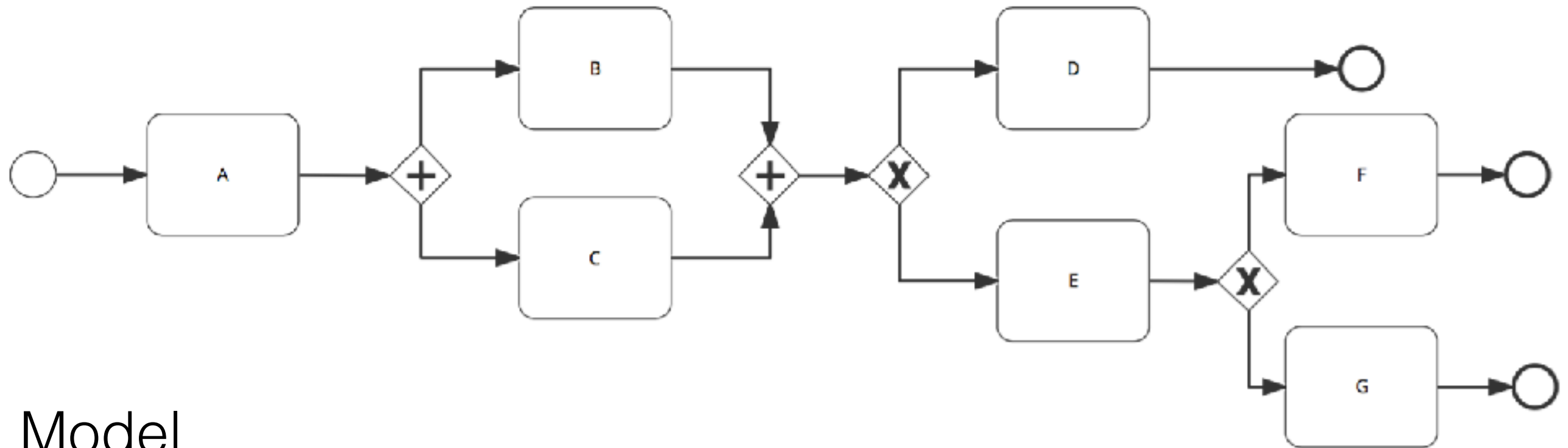
Event Databases

- Data sets based on news reports from a variety of international news sources coded using a coding framework
- The Global Database of Events, Language, and Tone (GDELT) available on Google Cloud Platform

Research Aim

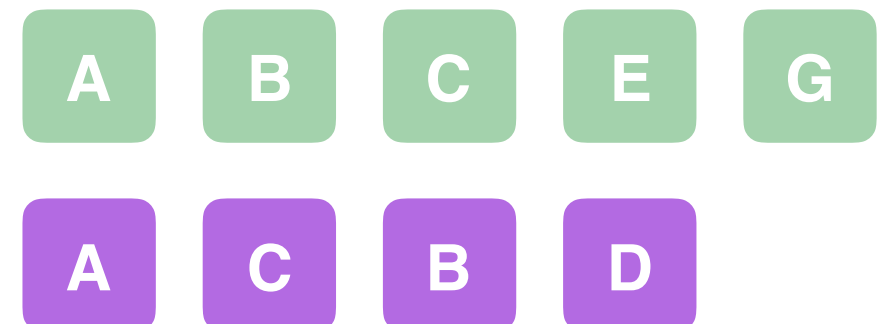
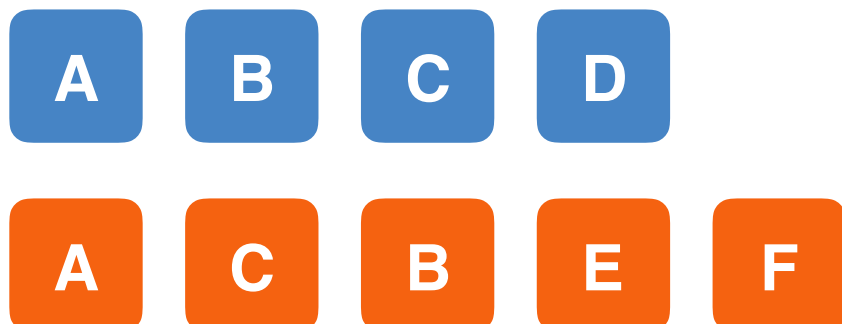
- Can we discover regularities in event occurrences?
- ...by adopting a process perspective?

Introduction to process mining



Model

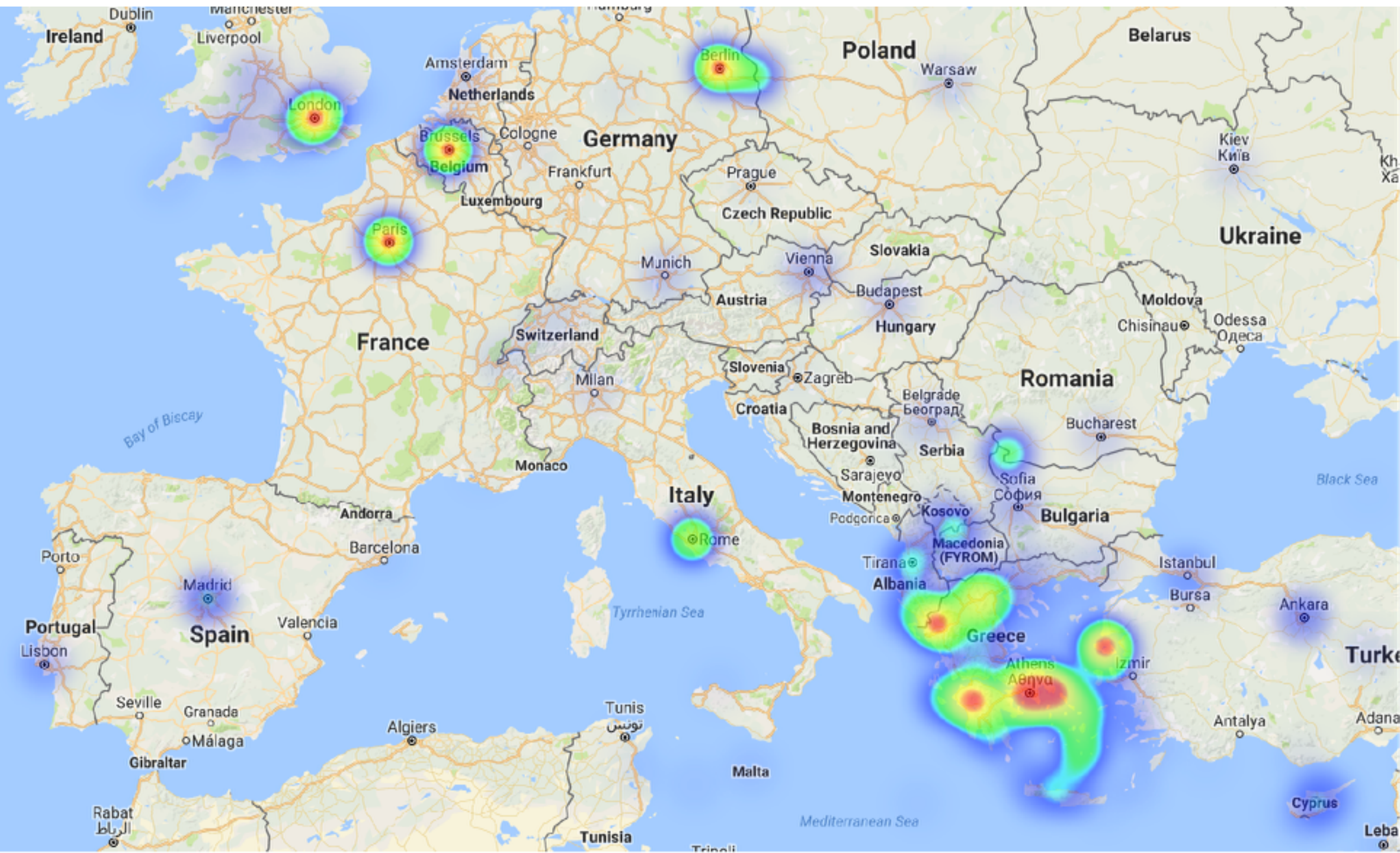
Traces



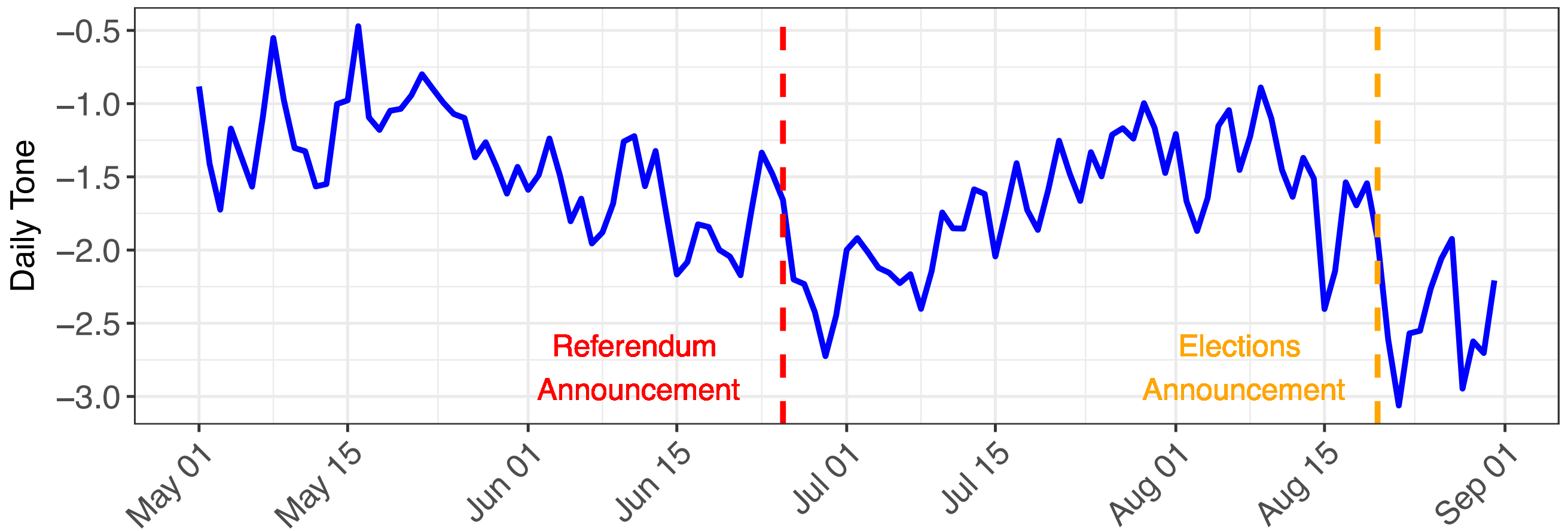
A Case Study



(Ordinary) Event Analytics



(Ordinary) Event Analytics



(Ordinary) Event Analytics



Issues of ordinary event analytics

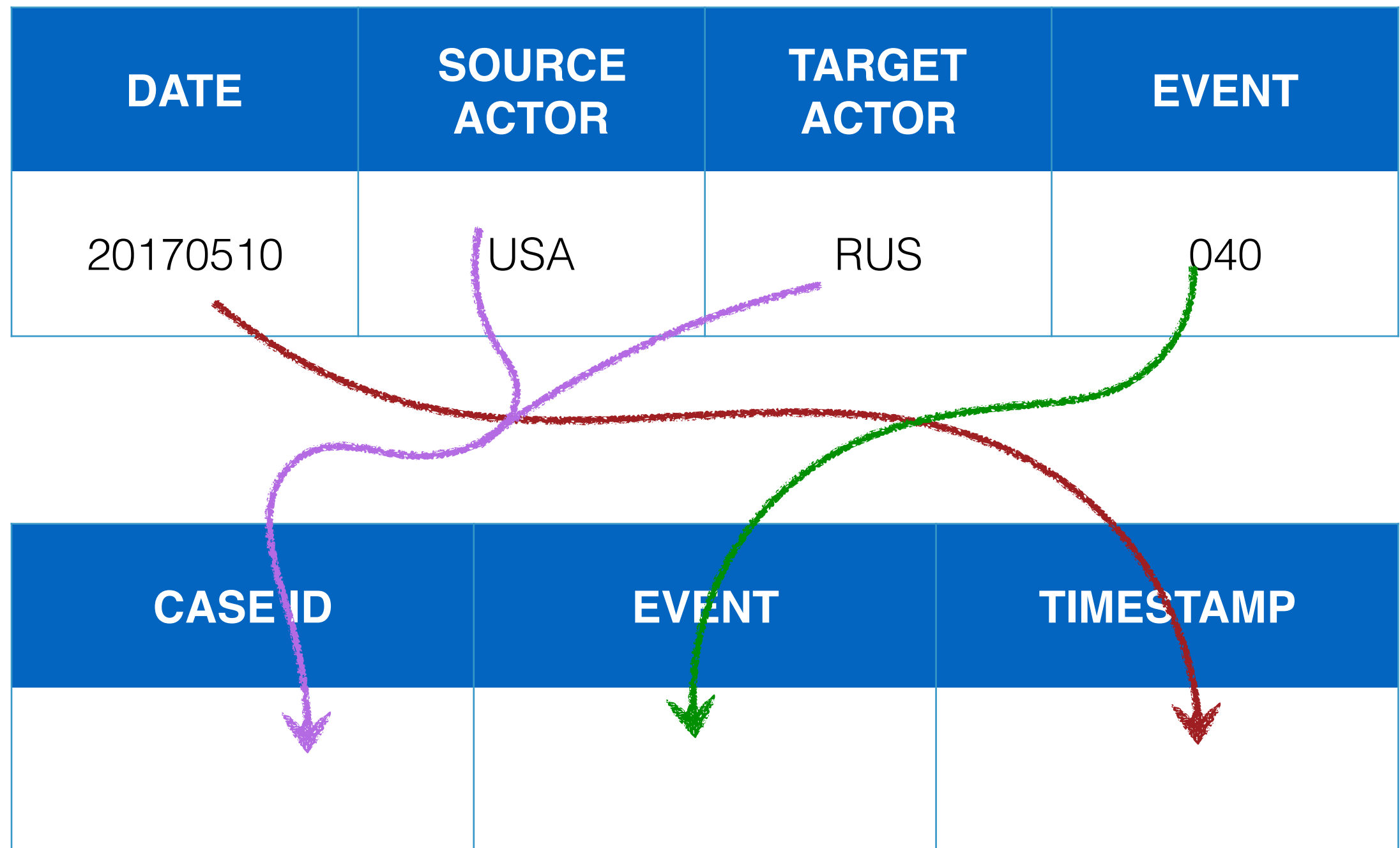
- Events are considered fragmentary, "rambling" elements of the story (*in aggregation techniques*)
- Focus on the pure sequentiality of events' occurrences (*in flow diagrams*)
- ...but real-world news events often do not follow a linear movement that is implied by a sequence

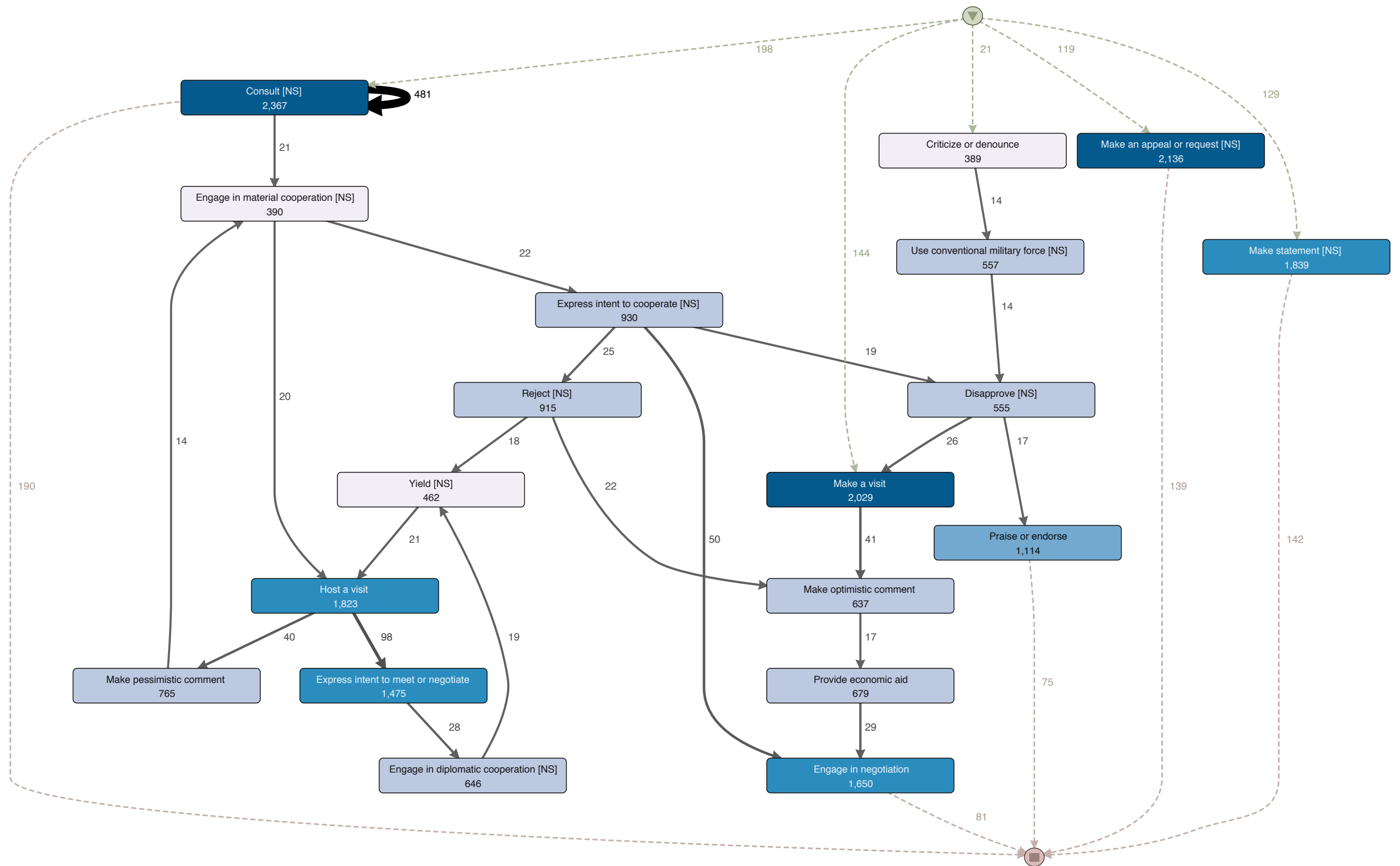
Transformation proposals

DATE	SOURCE ACTOR	TARGET ACTOR	EVENT
20170510	USA	RUS	040

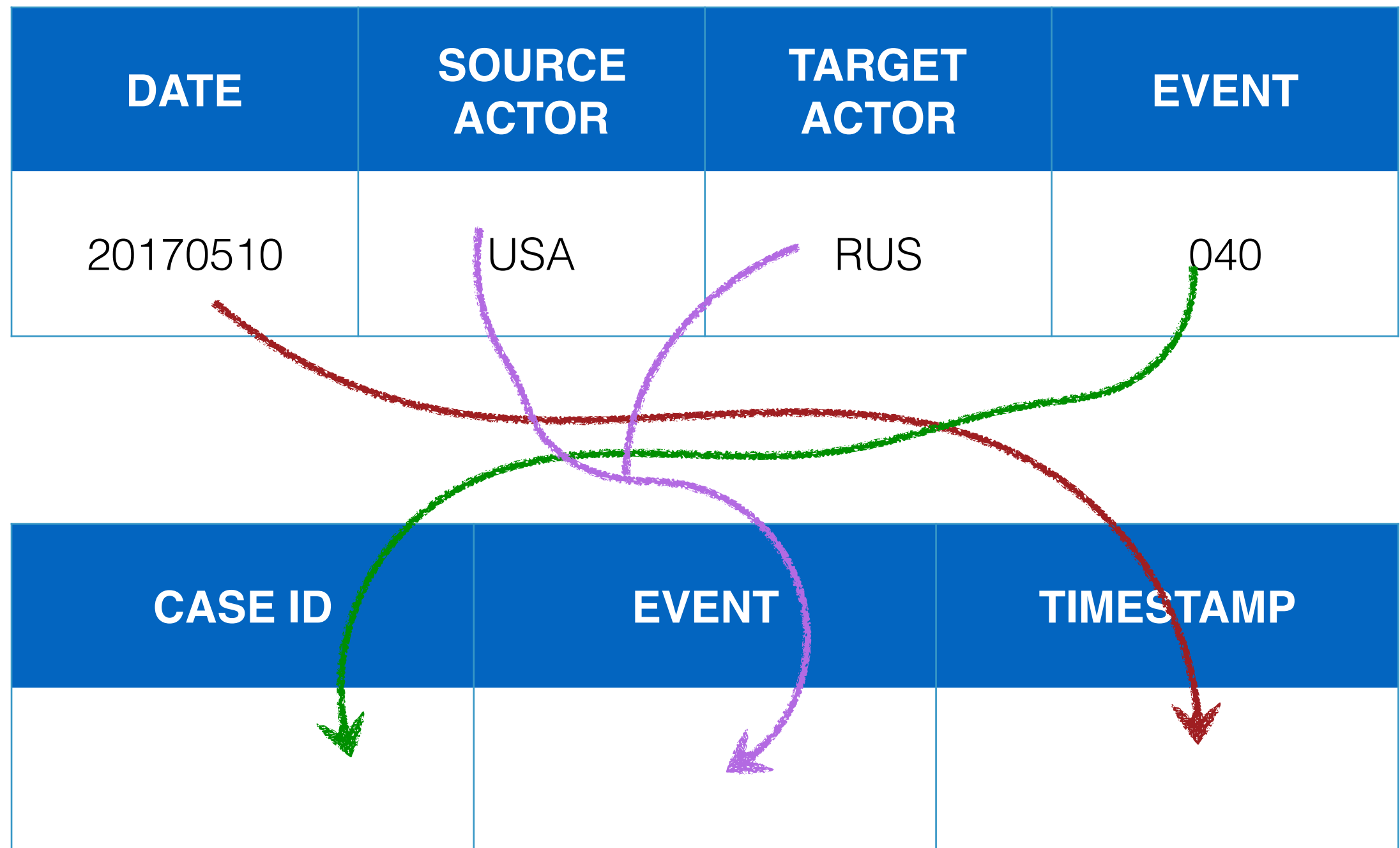
CASE ID	EVENT	TIMESTAMP

Transformation proposals



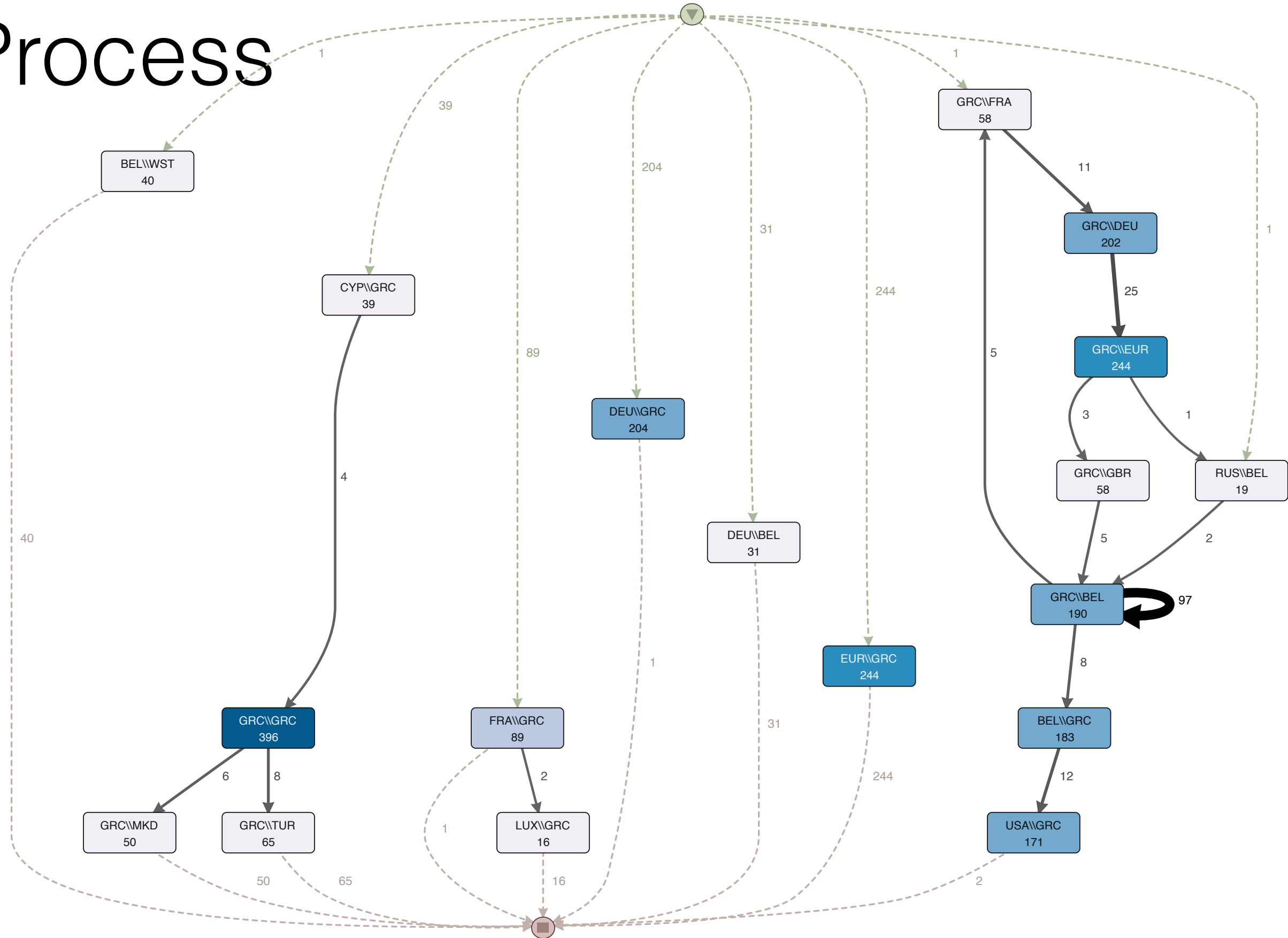


Transformation proposals





The “Consult” Process



Contribution

- Enable a process perspective for the (otherwise disjointed) events
- An additional analytic capability

Η περίπτωση μιας
δημοτικής υπηρεσίας:
Πολεοδομία

APPROVED

BUILDING PERMIT

This card must be kept posted in a conspicuous place on site of

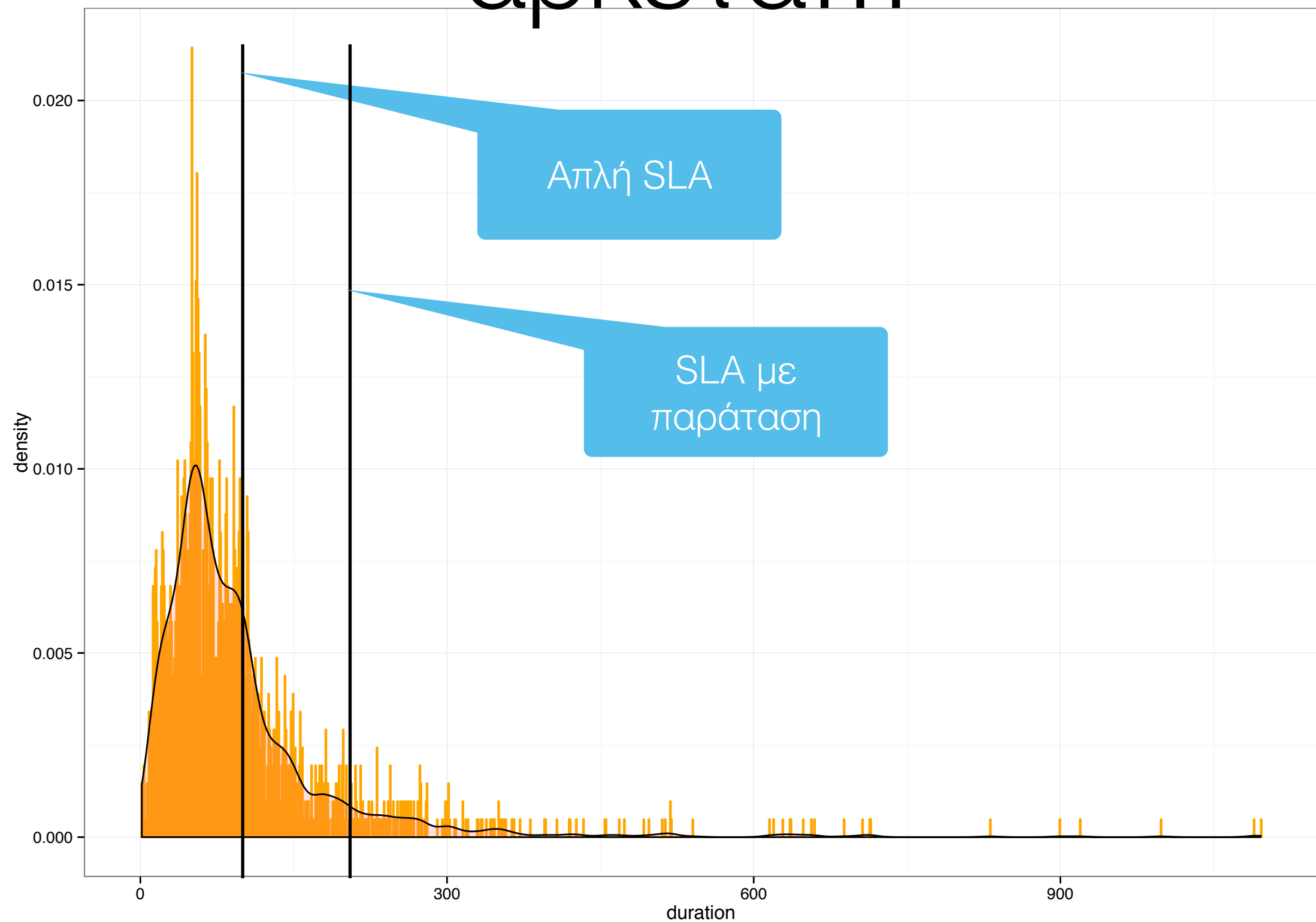
Πλαίσιο Εφαρμογής

- Ανασχετά δεδομένα 5 Δήμων της Ολλανδίας (BPIC'15)
- 262,628 γεγονότα που αντιστοιχούν σε that 5,649 αιτήσεις
- Μία περίοδο περίπου πέντε ετών (τέλη 2009 - αρχές 2015)

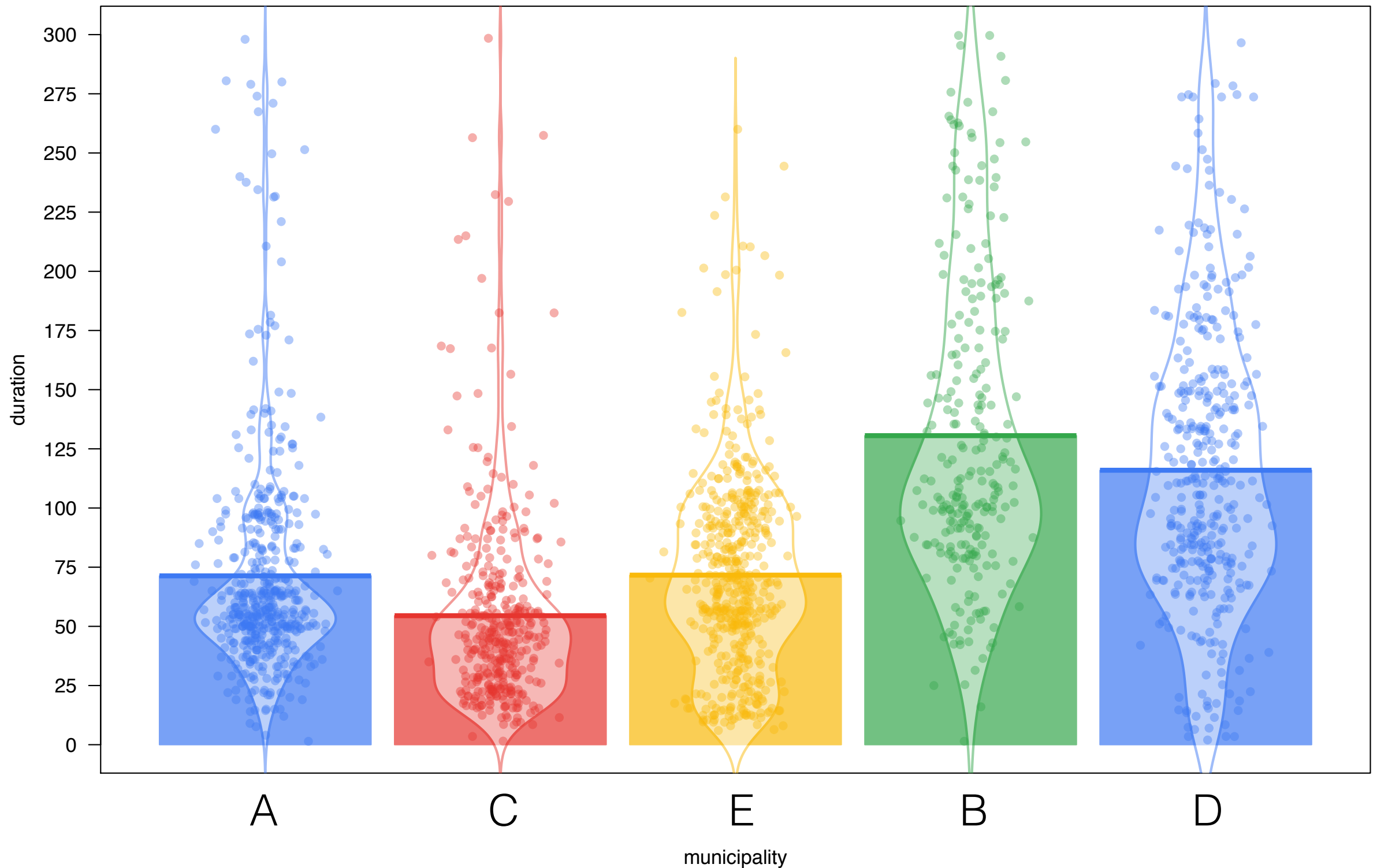
Πως μοιάζουν τα δεδομένα;

Case.ID	Timestamp	Activity	...
2771451	10/6/10	register submission date request	
2771451	10/19/10	enter senddate procedure confirmation	
2783345	12/1/10 10:32	applicant is stakeholder	
2783345	12/1/10 10:38	ask stakeholders views	

Οι επιδόσεις διαφέρουν αρκετά...



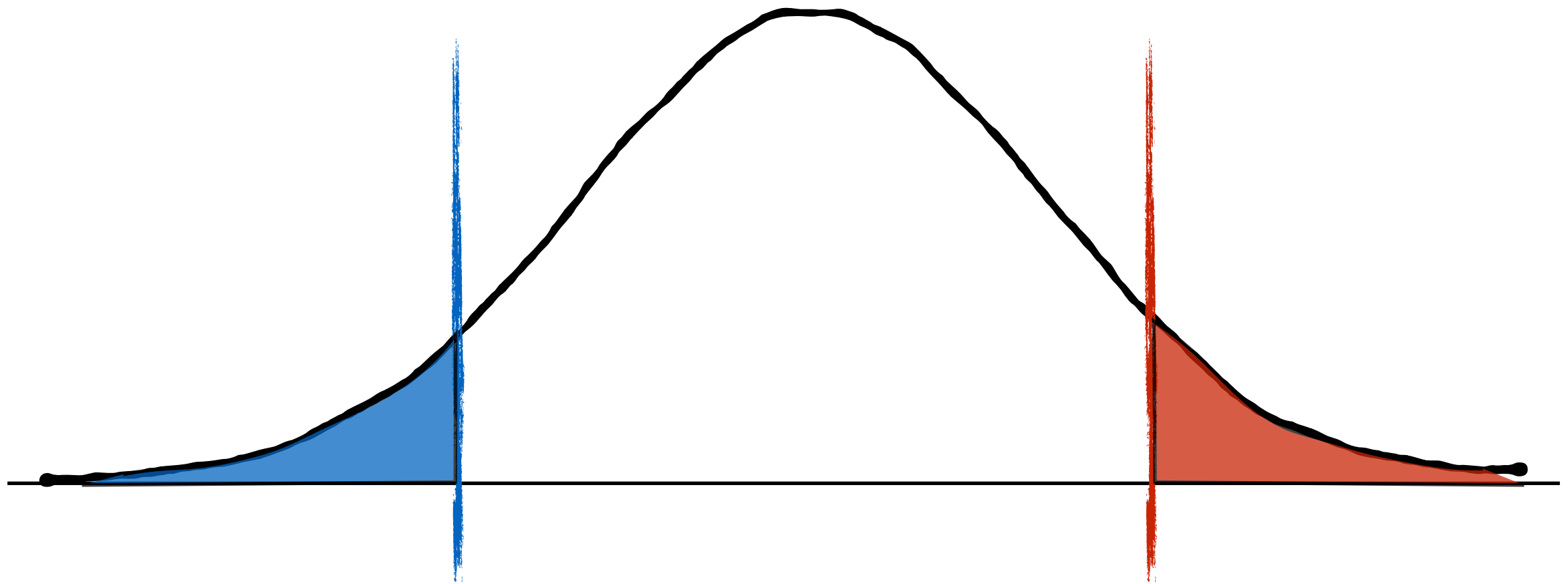
Ίδιος νόμος, Διαφορετικές επιδόσεις



Γιατί πρέπει να απαντάμε έγκαιρα;

- Διαφυγόντα κέρδη (πρόστιμα)
- Άδειες σε καθεστώς υποχρεωτικής χορήγησης
- It's the citizen, stupid!

Η βασική ιδέα



Μάθηση από τις θετικές αποκλίσεις

Θετική Απόκλιση

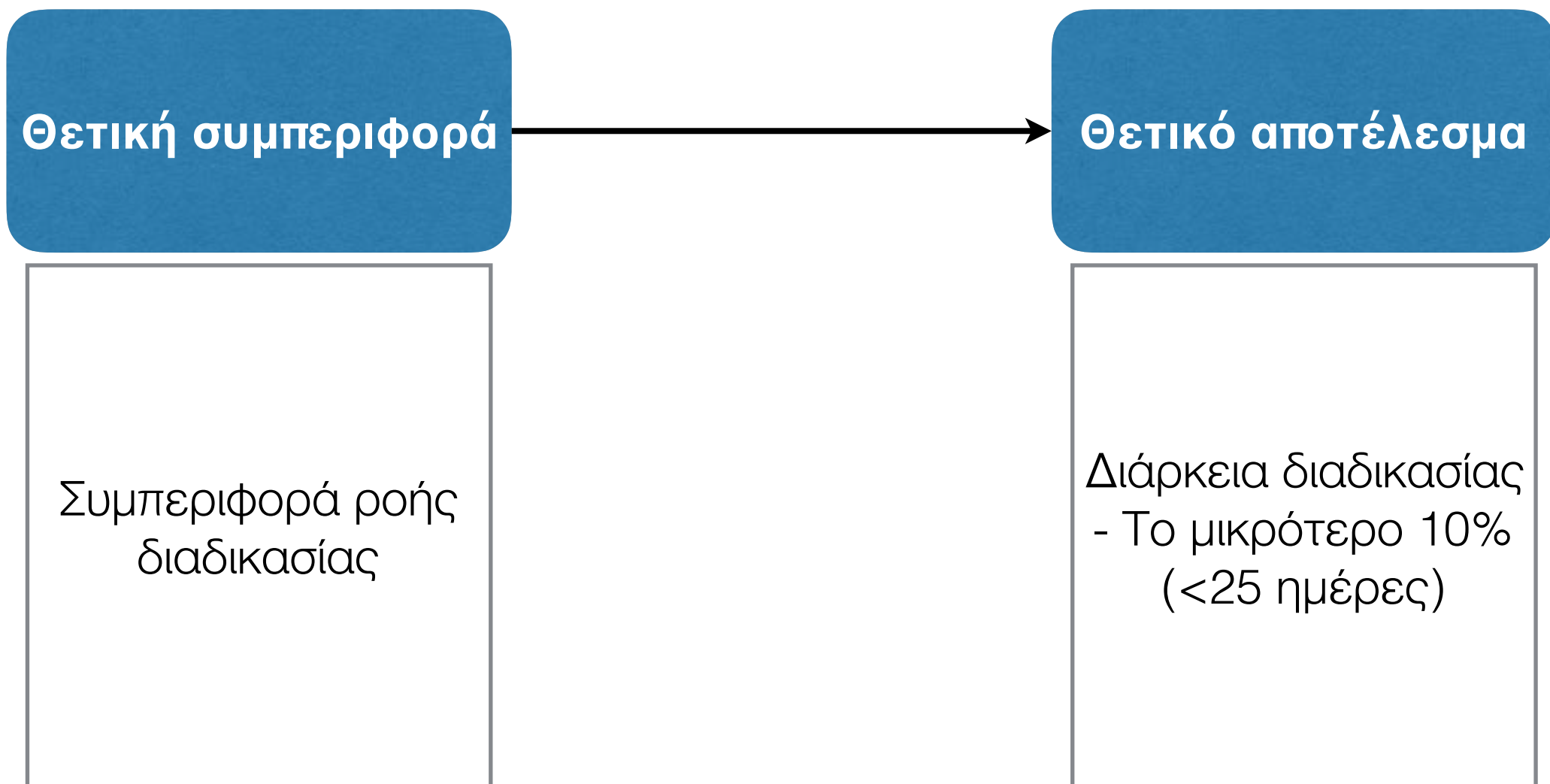
- Πρώιμη εμφάνιση στο πεδίο της επιστήμης της διατροφής
- Οι κοινότητες γνωρίζουν τις λύσεις. Περιέχουν τους ειδήμονες για την επίλυση των προβλημάτων.
- Προτείνει βελτιώσεις μέσα από την ανάλυση φαινομένων θετικών αποκλίσεων



Ερευνητικές Υποθέσεις

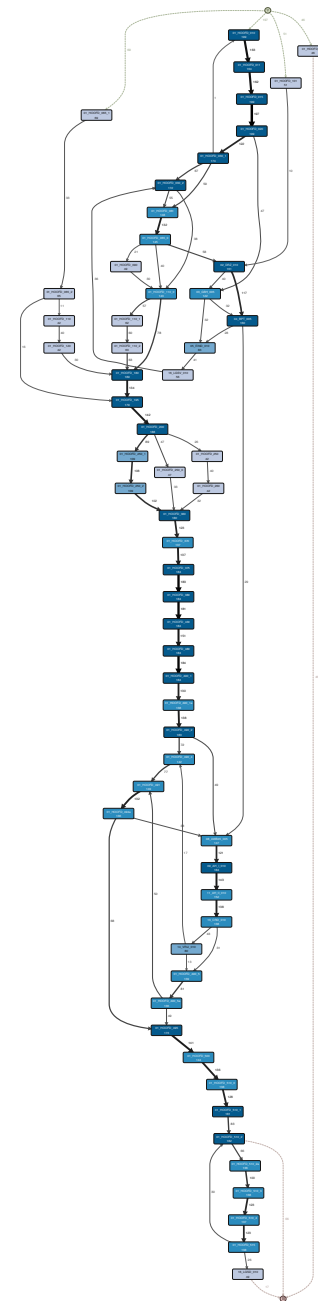
- Έχουν οι θετικά αποκλίνουσες περιπτώσεις διαφορετική ροή εξέλιξης;
- Μπορούν αυτές οι διαφορές στην ροή να προτείνουν εργασίες οι οποίες, εφόσον ανατεθούν σε τρίτους, μπορούν να βελτιώσουν τις επιδόσεις;

Το αντικείμενο της θετικής απόκλισης

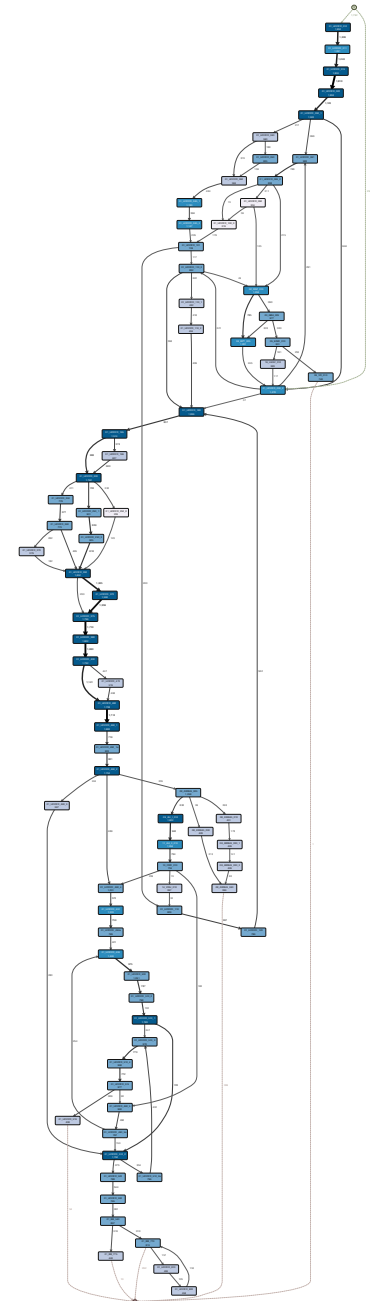


Έχουν οι θετικά αποκλίνουσες περιπτώσεις διαφορετική ροή εξέλιξης;

- Ανακάλυψη δύο μοντέλων διαδικασιών
- Φόρμουλα ομοιότητας μοντέλων
- Τεράστιες διαφορές (>170 εργασίες)



Θετικές



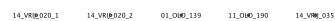
Κανονικές

Μπορούν οι διαφορές στην ροή να προτείνουν εργασίες προς ανάθεση σε τρίτους;

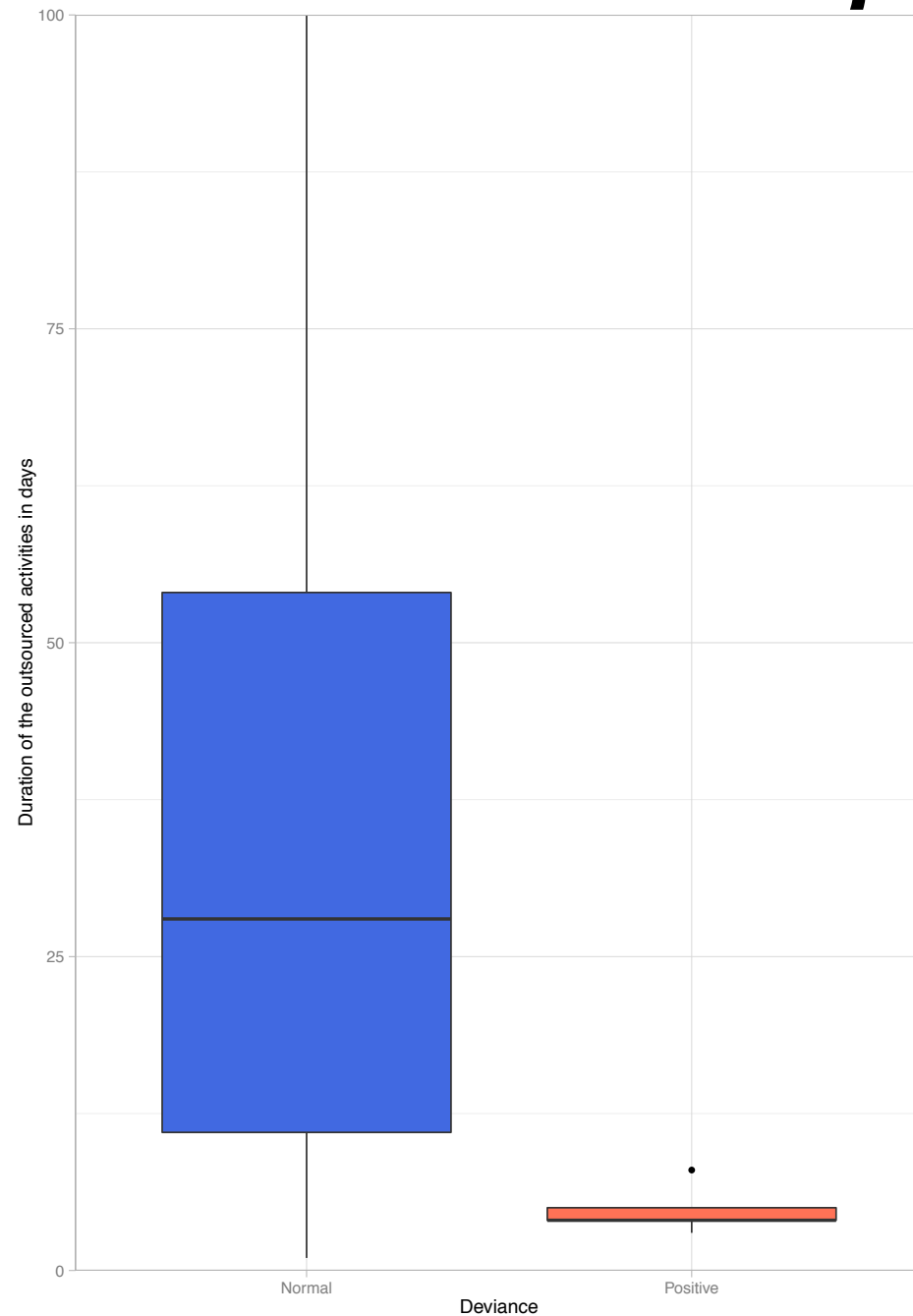
- Υποθέτουμε δύο χαρακτηριστικά που πρέπει να διαθέτουν οι προς ανάθεση εργασίες:
 - Οι μεταβιβάσεις/ αλληλεπιδράσεις μεταξύ υπεργολάβων και δήμων πρέπει να είναι οι ελάχιστες δυνατές
 - Οι προς ανάθεση εργασίες πρέπει να εμφανίζονται συχνότερα στα περιστατικά με κανονική διάρκεια, παρά στα θετικά αποκλίνοντα.

Μεθοδολογία

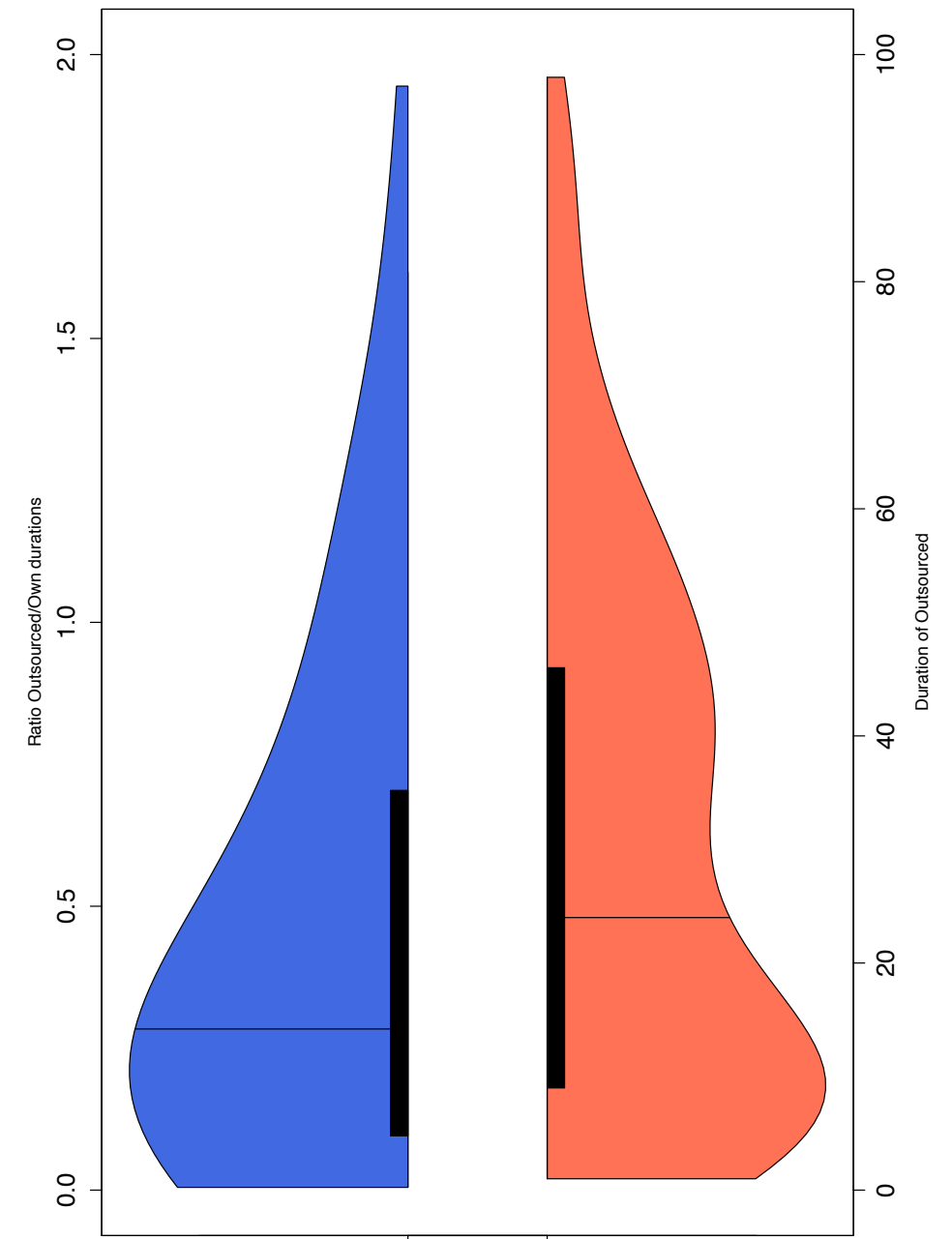
- Δημιουργία ενός δικτύου (κόμβοι οι εργασίες, βαροδοτημένες ακμές για τις συνδέσεις)
- Διαχωρισμός του δικτύου (*Spectral Clustering*)
- Προσαρμογή του μεγέθους των κόμβων ανάλογα με τις διαφορές των συχνοτήτων
- Διάταξη Backbone για απομόνωση γειτονιών εργασιών (*spanning subgraph*)
- Ποιοτική εκτίμηση



Αξιολόγηση



Άθροισμα διάρκειας των προς ανάθεση εργασιών για κανονικές και θετικά αποκλίνουσες περιπτώσεις



Αριστερά: Κλάσμα συνολικής διάρκειας των υπό ανάθεση εργασιών προς συνολική διάρκεια υπόλοιπων εργασιών
Δεξιά: Κατανομή της συνολικής διάρκειας των προς ανάθεση εργασιών



Ευχαριστώ!

Η Εξόρυξη Διαδικασιών ως Εργαλείο Αποφάσεων

Αναπληρωτής Καθηγητής Παύλος Δελιάς,

Τμήμα Λογιστικής & Χρηματοοικονομικής

pdelias@teikav.edu.gr, @PavlosDelias