

# VIS 2021

---

## Mining and Understanding Stories in Text Sequences with Narrative Visualization



Award for Innovative Narrative Visualization and Analysis Methodology



**Zeyu Li**, Teng Wang, Ruizhi Shi, Zhaohui Li, Jiawan Zhang  
Tianjin University

# Agenda

- Overview
- Solutions
  - Q1
    - a classification system
    - a system used for verifying and exploring the classification
  - Q2: a narrative visualization creation system
  - Q3: a map with trajectory
  - Q4: summarized by human
- Conclusion

# Overview: data

- Microblog messages
  - date
  - author
  - message
  - location(optional)

text sequences with timestamp

- Call center data
  - date,
  - message
  - location

spatiotemporal data

# Overview: questions

- Organize, understand and summarize the content of messages from a retrospective perspective
  - Q1: classify messages and identify the feature of each class
  - Q2: analyze the evolution of the level of the risk
- Compare the differences between the retrospective analysis and real-time analysis in terms of decision making and visualization solution
  - Q3: determine a dispatch location for first responders
  - Q4: compare the previous and current solution

# Overview: questions

**Q1:** classify messages and identify the feature of each class

**Q2:** analyze the evolution of the level of the risk

**Q3:** determine a dispatch location for first responders

**Q4:** compare the previous and current solution

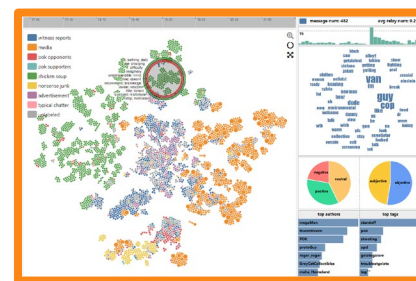
#1



classified  
messages



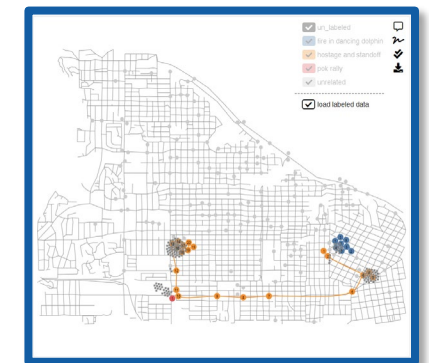
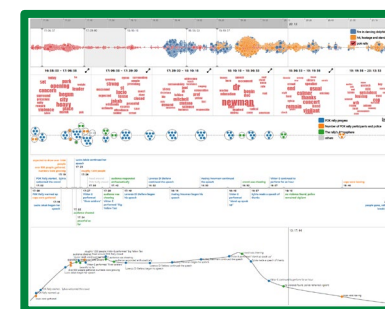
#2



messages of  
three main  
events



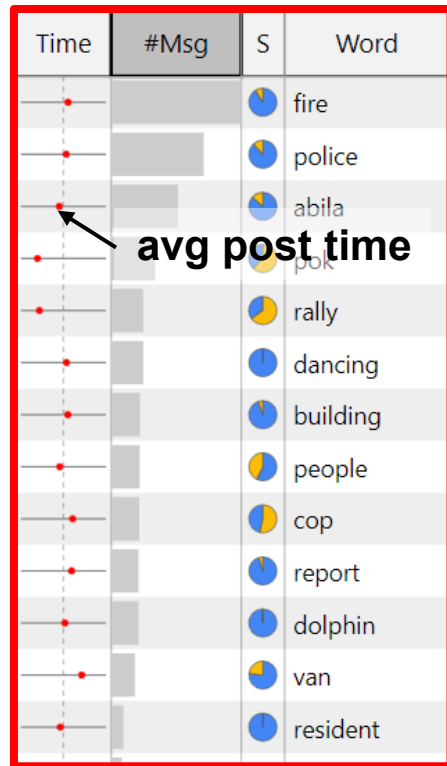
#3



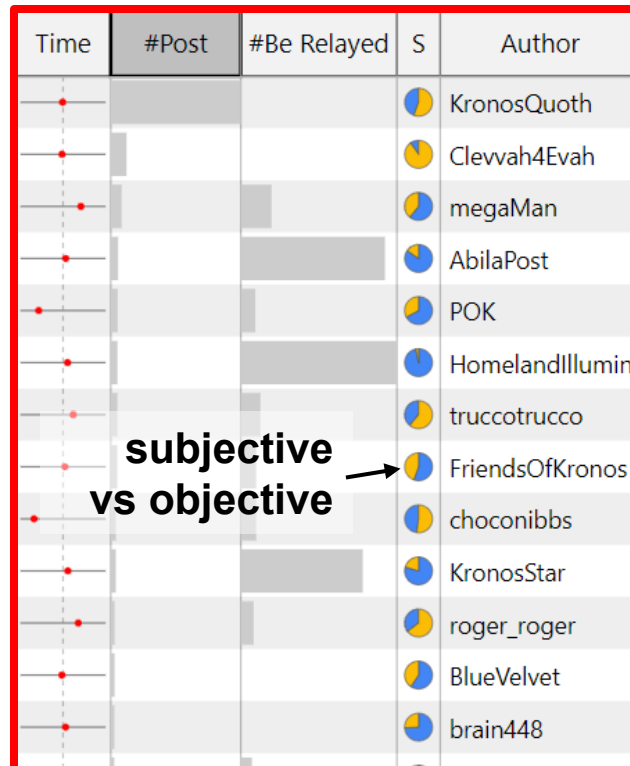
Q1: classify messages according to their content and to identify the features of each class

# First system: classify messages

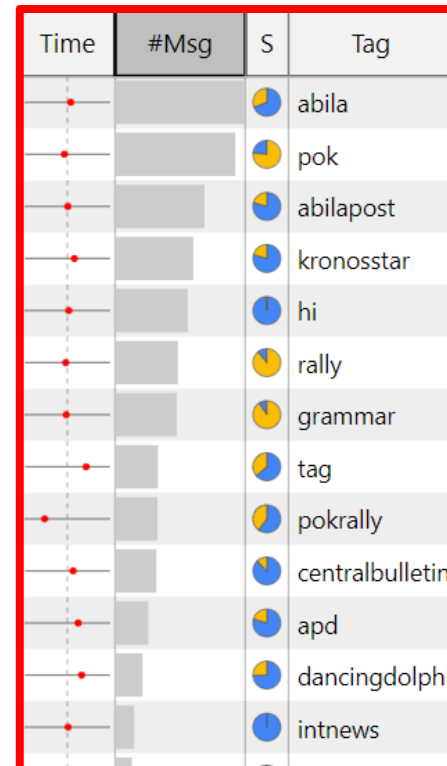
top keywords



top authors



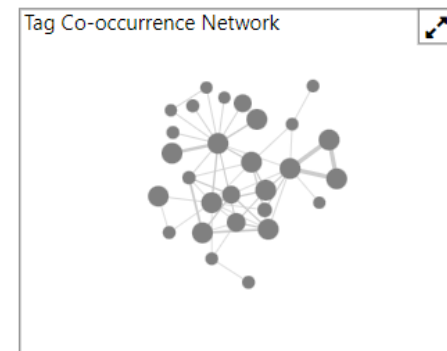
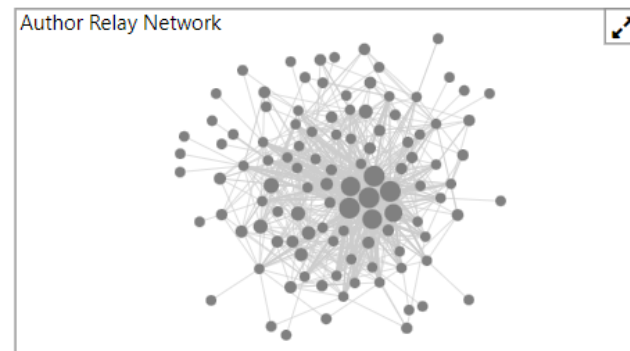
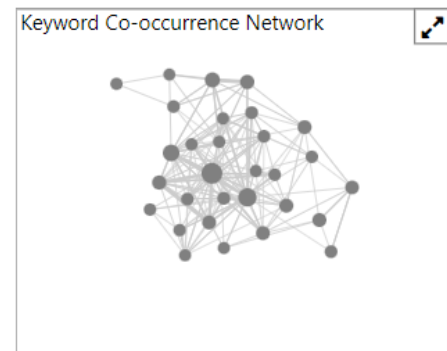
top tags



messages of an instance

S	C	Message	Q	□	□
●	●	Thinking should become your capital asset, no matter whatever ups and downs you come across in your life. #POKRally #HI			□
●	●	RT @MindOfKronos Ugh, these POKers are anooying! #POK #gowaway			□
●	●	RT @CentralBulletin POK rally today in Abila City Park concerns of violence and heavy police presence #Abila #CentralBulletin			□
●	●	RT @CentralBulletin POK rally today in Abila City Park concerns of violence and heavy police presence #Abila #CentralBulletin			□
●	●	RT @ourcountryourrights Sylvia Marek opens with welcoming remarks. #POK			□
●	●	POK rally to start in Abila City Park. POK leader Sylvia Marek to open with a speech? #IntNews			□
●	●	If you can dream it, you can achieve it. #KronosStar #POKRally			□
●	●	POK trying to give an aura of cred - while kidnapping, extorting			□
●	●	RT @AbilaPost Special guest Dr. Audrey McConnel Newman will be speaking along with Lucio Jakab and Professor Lorenzo Di Stefano			□
●	●	Only put off until tomorrow what you are willing to die having left undone. #POKRally #HI			□
●	●	Prof Di Stefano should lose his job #POK #terrorsympathizer			□

subjective or objective



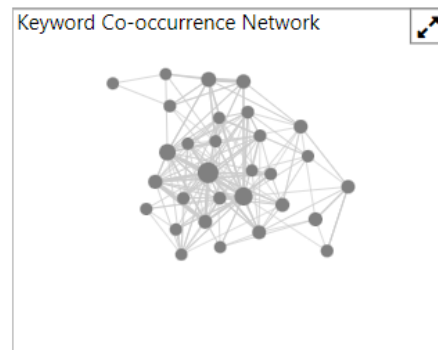
classification tree

● root 0

# First system: classify messages

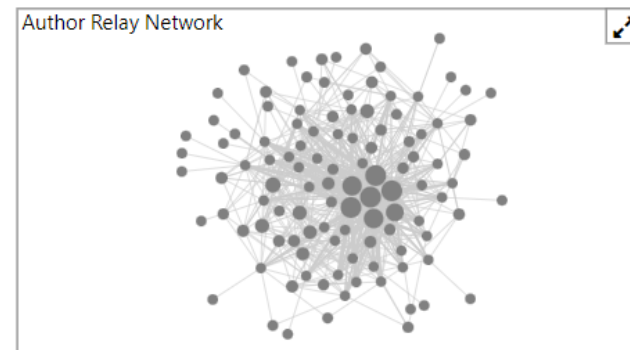
top keywords

Time	#Msg	S	Word
		🟦	fire
		🟦	police
		🟦	abila
		🟡	pok
		🟡	rally
		🟦	dancing
		🟦	building
		🟡	people
		🟡	cop
		🟦	report
		🟦	dolphin
		🟡	van
		🟦	resident



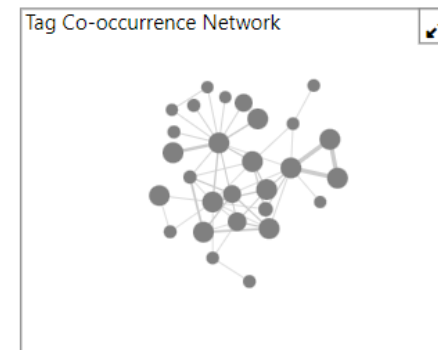
top authors

Time	#Post	#Be Relayed	S	Author
			🟡	KronosQuoth
			🟡	Clevah4Evah
			🟦	megaMan
			🟦	AbilaPost
			🟡	POK
			🟦	HomelandIllumina
			🟡	truccotrucco
			🟡	FriendsOfKronos
			🟡	choconibbs
			🟦	KronosStar
			🟡	roger_roger
			🟡	BlueVelvet
			🟡	brain448



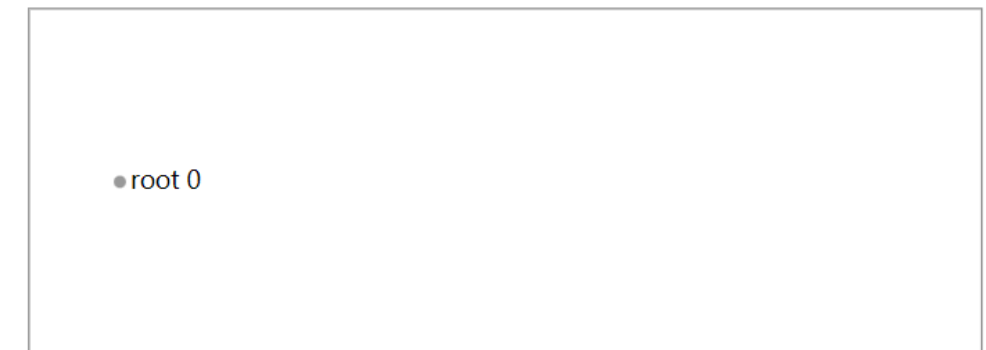
top tags

Time	#Msg	S	Tag
		🟦	abila
		🟡	pok
		🟦	abilapost
		🟦	kronosstar
		🟦	hi
		🟡	rally
		🟡	grammar
		🟦	tag
		🟡	pokrally
		🟦	centralbulletin
		🟦	apd
		🟡	dancingdolphi
		🟦	intnews



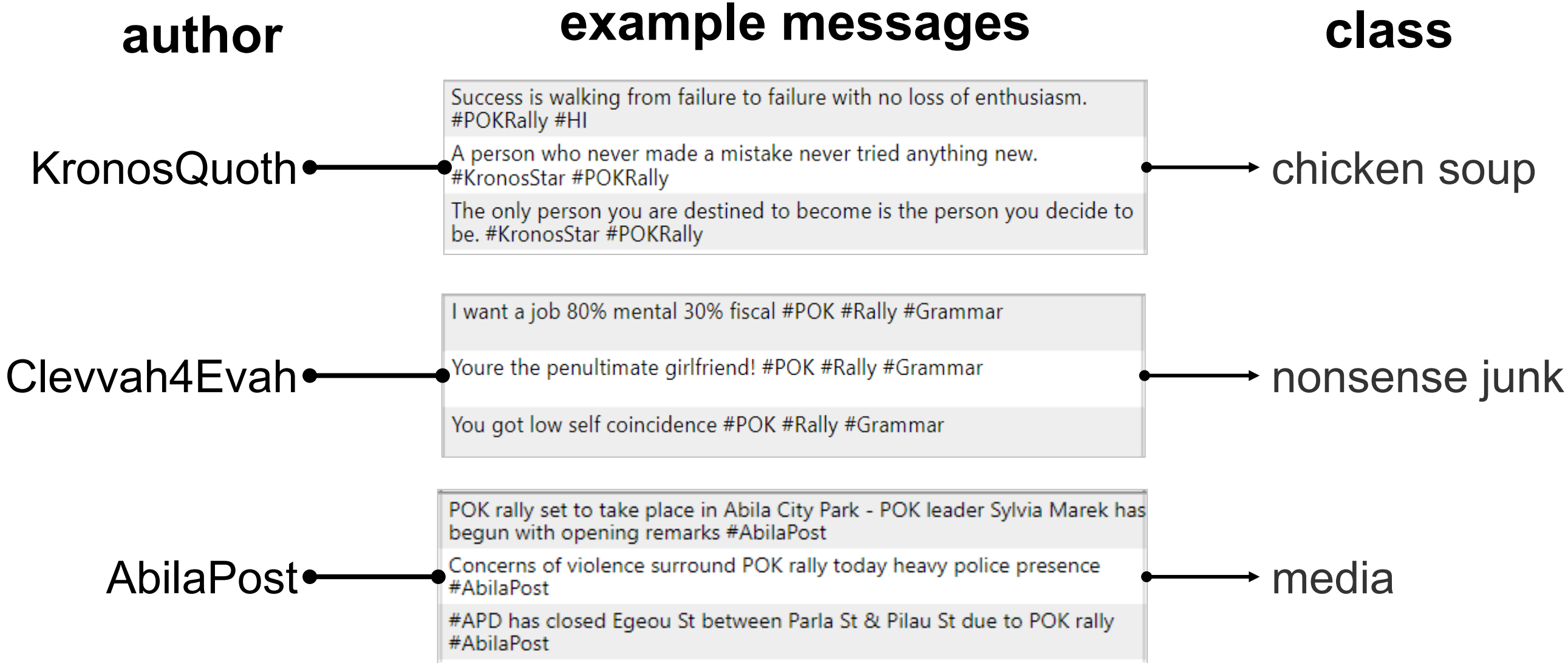
messages of an instance

S	C	author: megaMan	🔍	☐	☐
🟦	🟡	Run - don't walk across driveway into Gelato #almostcreamedbyvan			☐
🟡	🟡	Yeah van -- can't exit the backway at Gelato. And now the cops got him!			☐
🟡	🟡	Better than expected - cops block van at Gelato. Moving to get a better view.			☐
🟦	🟡	So van tried a t-turn to exit; cop blocked him off; van and cop doors open #troubleatgelato			☐
🟦	🟡	shot! moving behind mailbox peeking #troubleatgelato			☐
🟡	🟡	@ben bullets can't go thru mailbox! #thanks4nothing			☐
🟦	🟡	Black van faces off two cop cars, shots fired, people undercover, not moving until its under control #troubleatgelato			☐
🟦	🟡	@ben @george Can't see whos in the van yet, one cop			☐
🟡	🟡	cop down? #troubleatgelato			☐
🟦	🟡	Still can't tell - keeping my head down #troubleatgelato			☐
🟦	🟡	van guys behind their doors, cops too #troubleatgelato			☐

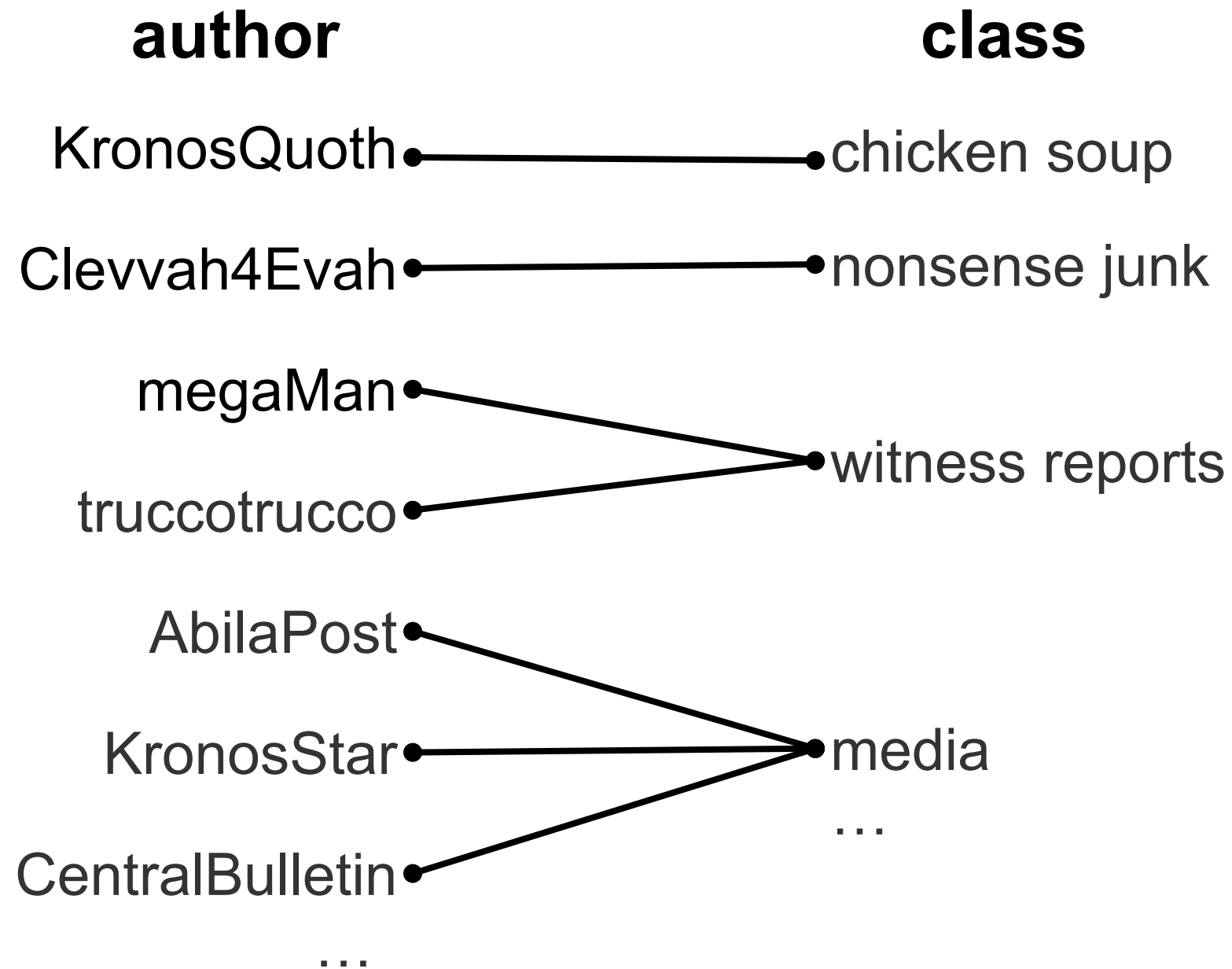




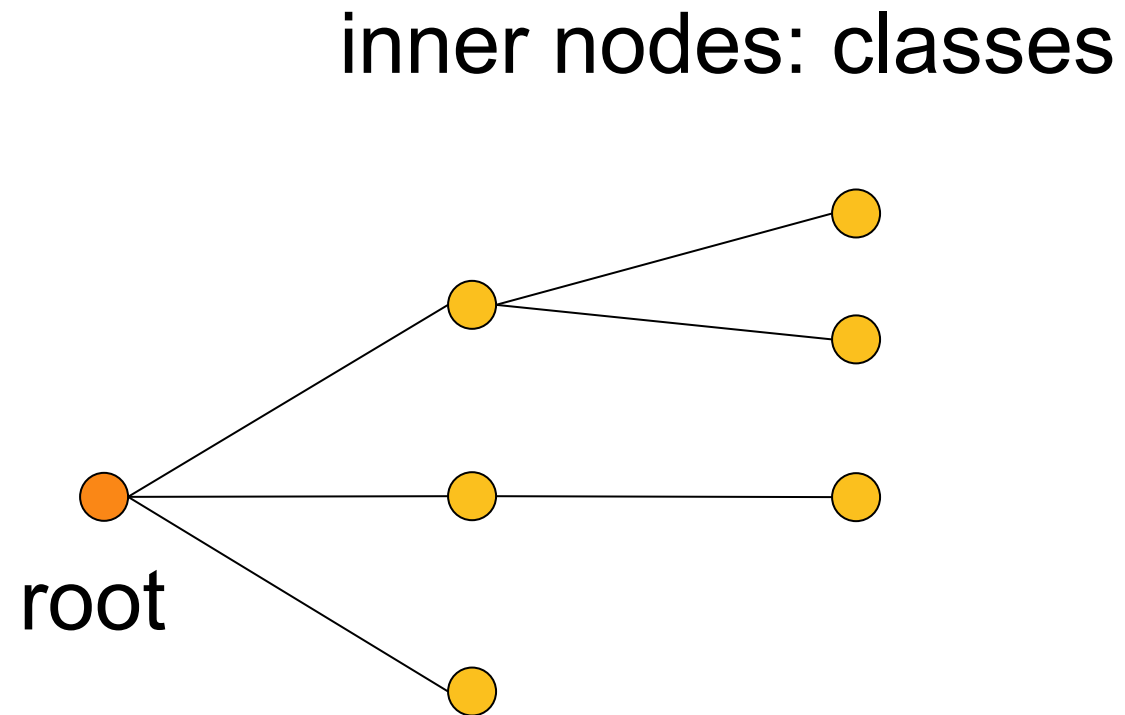
# Mapping between authors and classes



# Mapping between authors and classes



# Build a classification tree



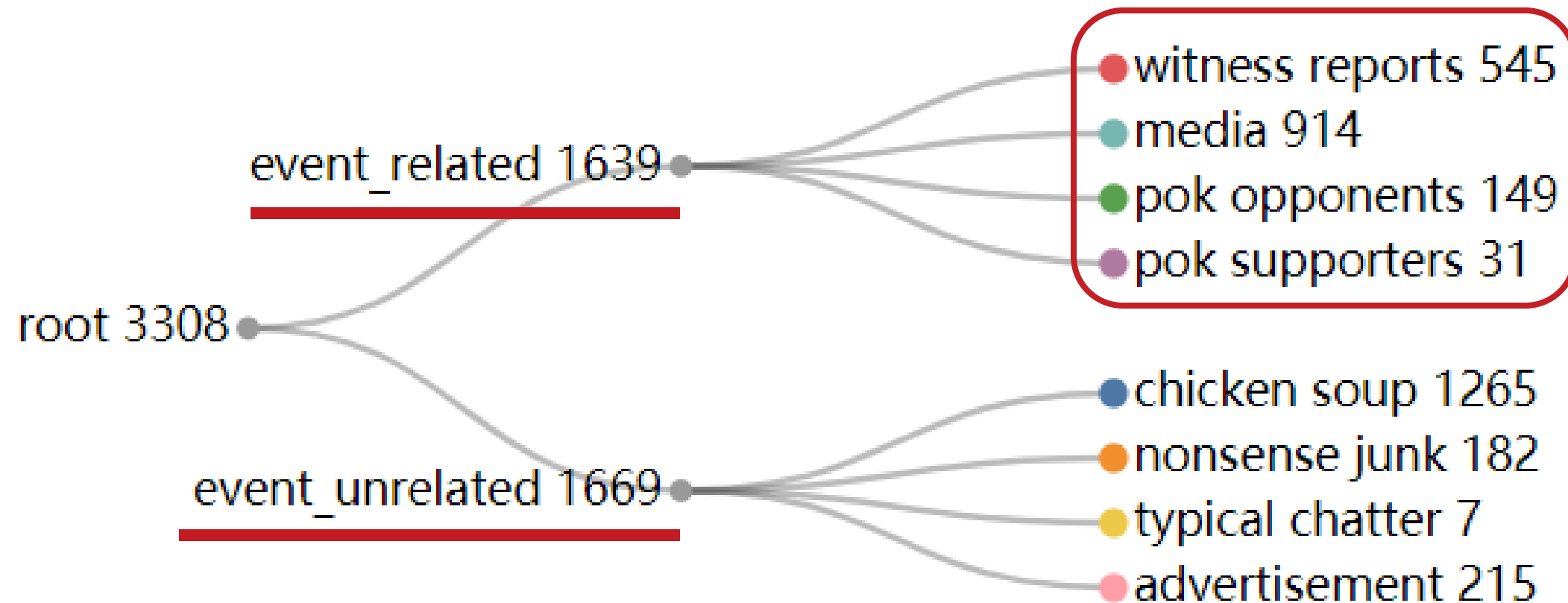
authors with messages



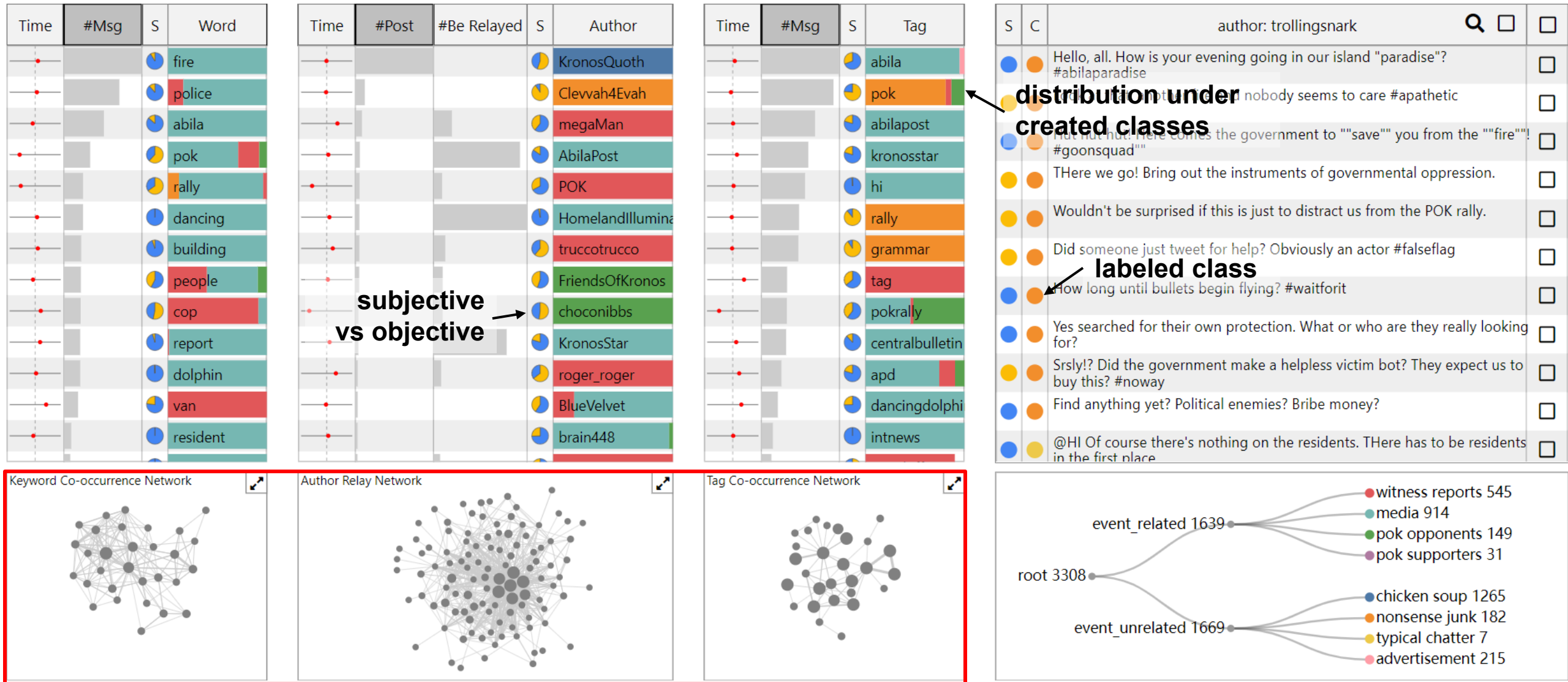
1. create a new class
2. assign an author' messages to an class

# A built classification tree

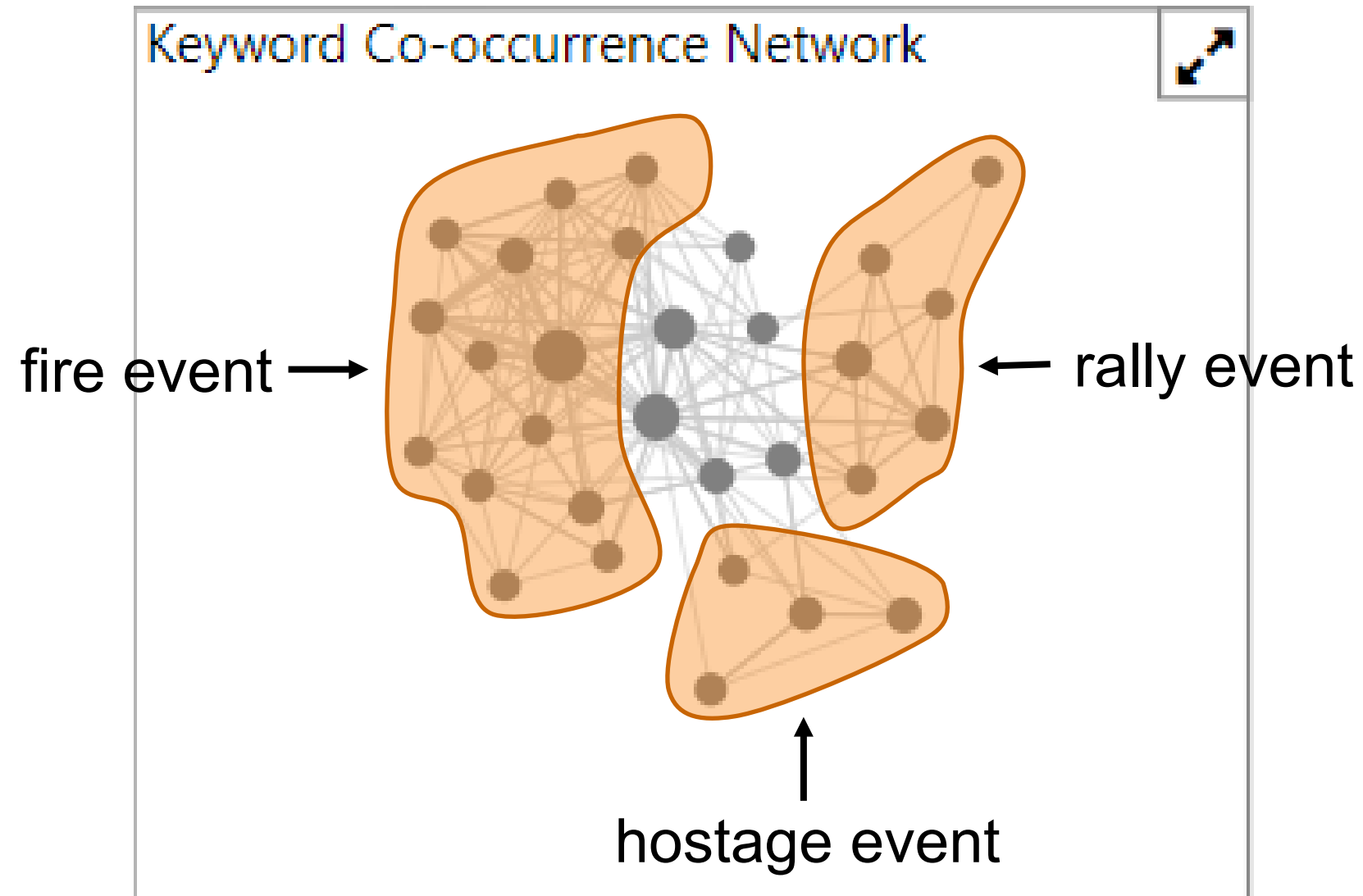
After labeling the top 30 authors, 95% of messages are assigned.



# Network of instances

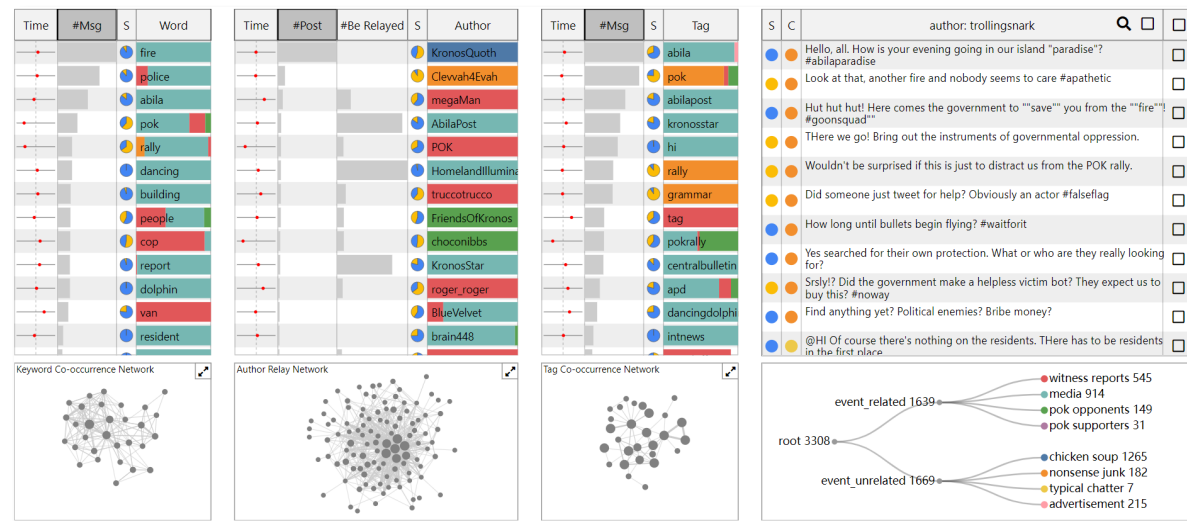


# Network of instances: three main events



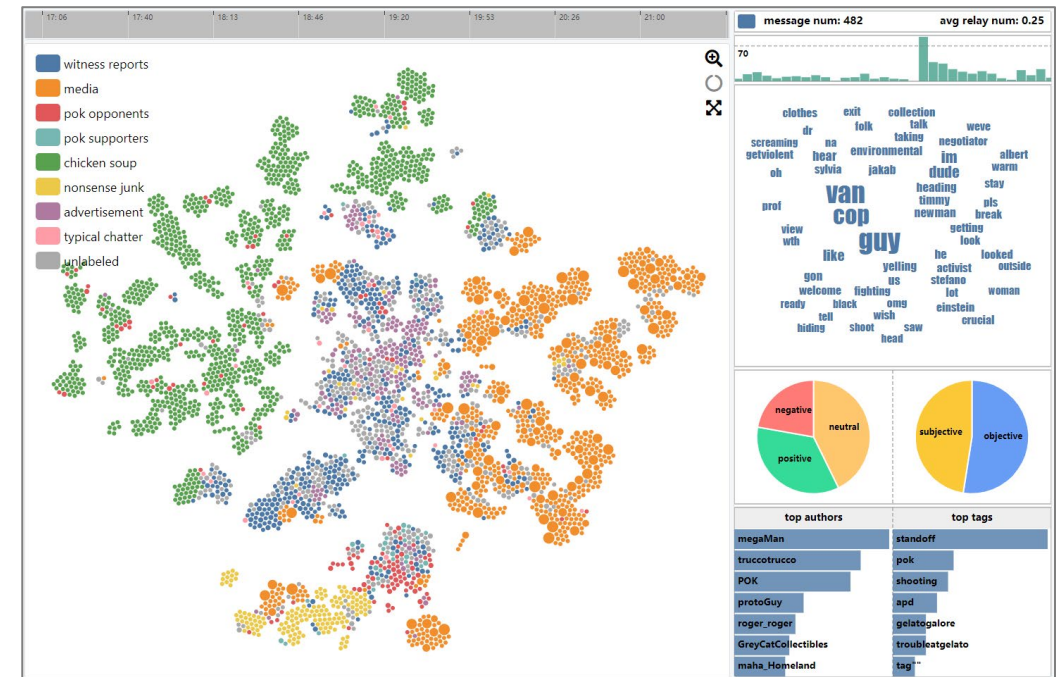
# Second system

## the first system



classified messages

## the second system



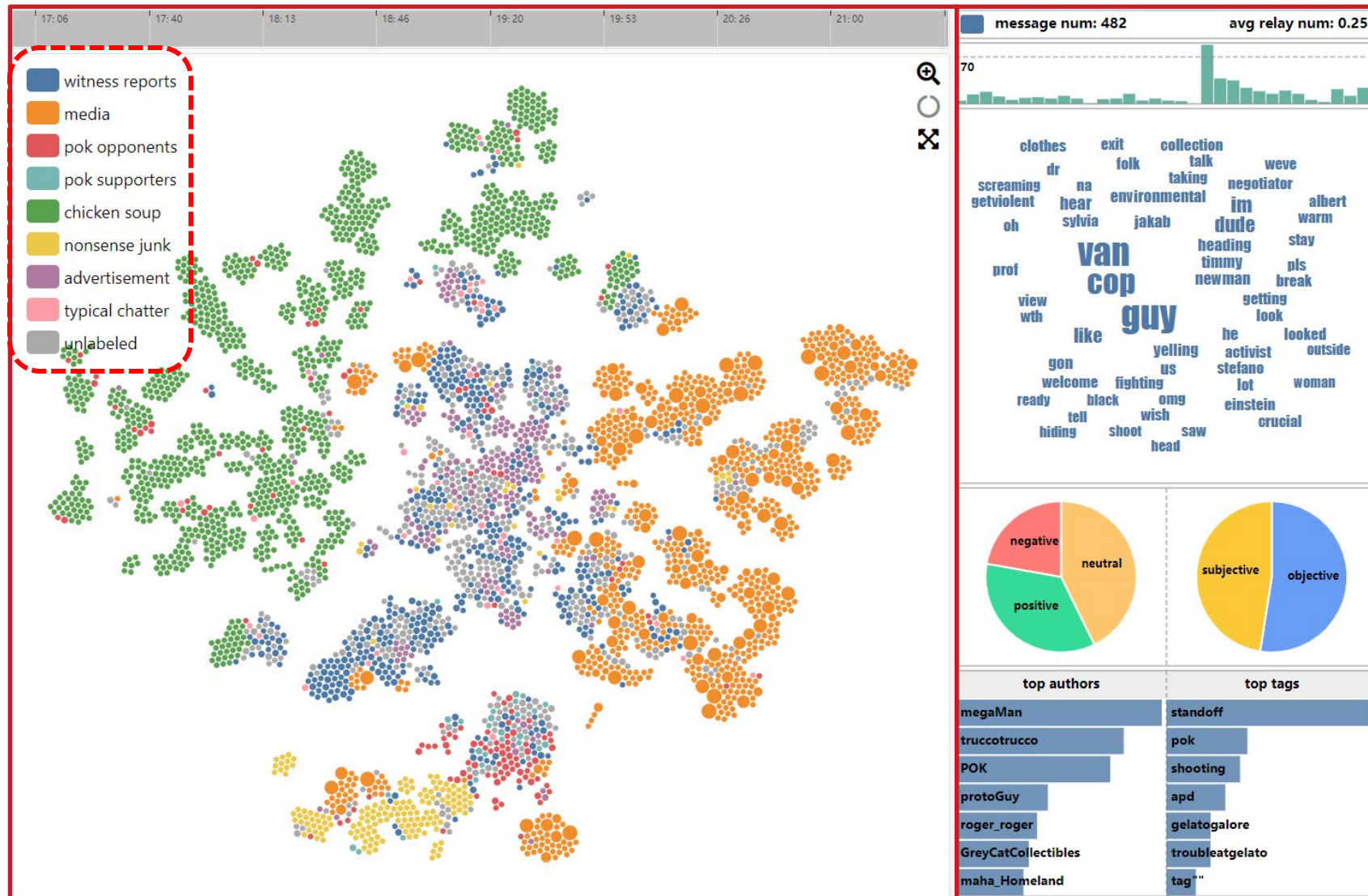
- check the reliability of the classification
- identify features of each class



# Second system

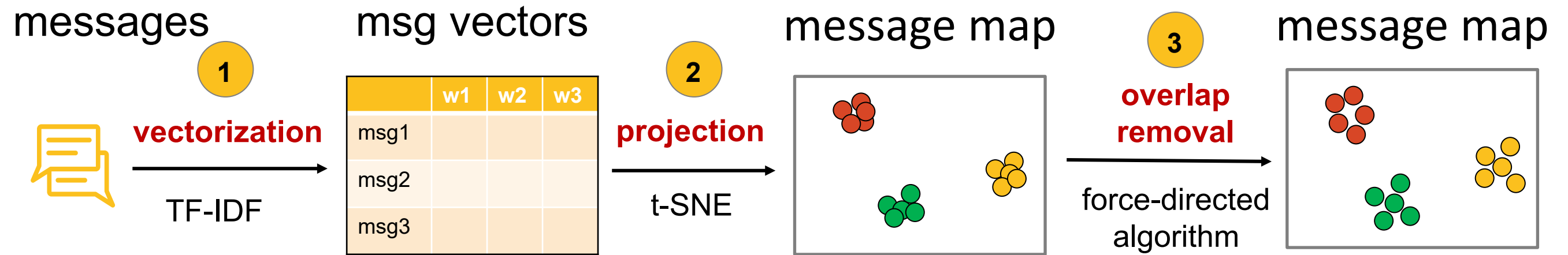
message map

feature checker



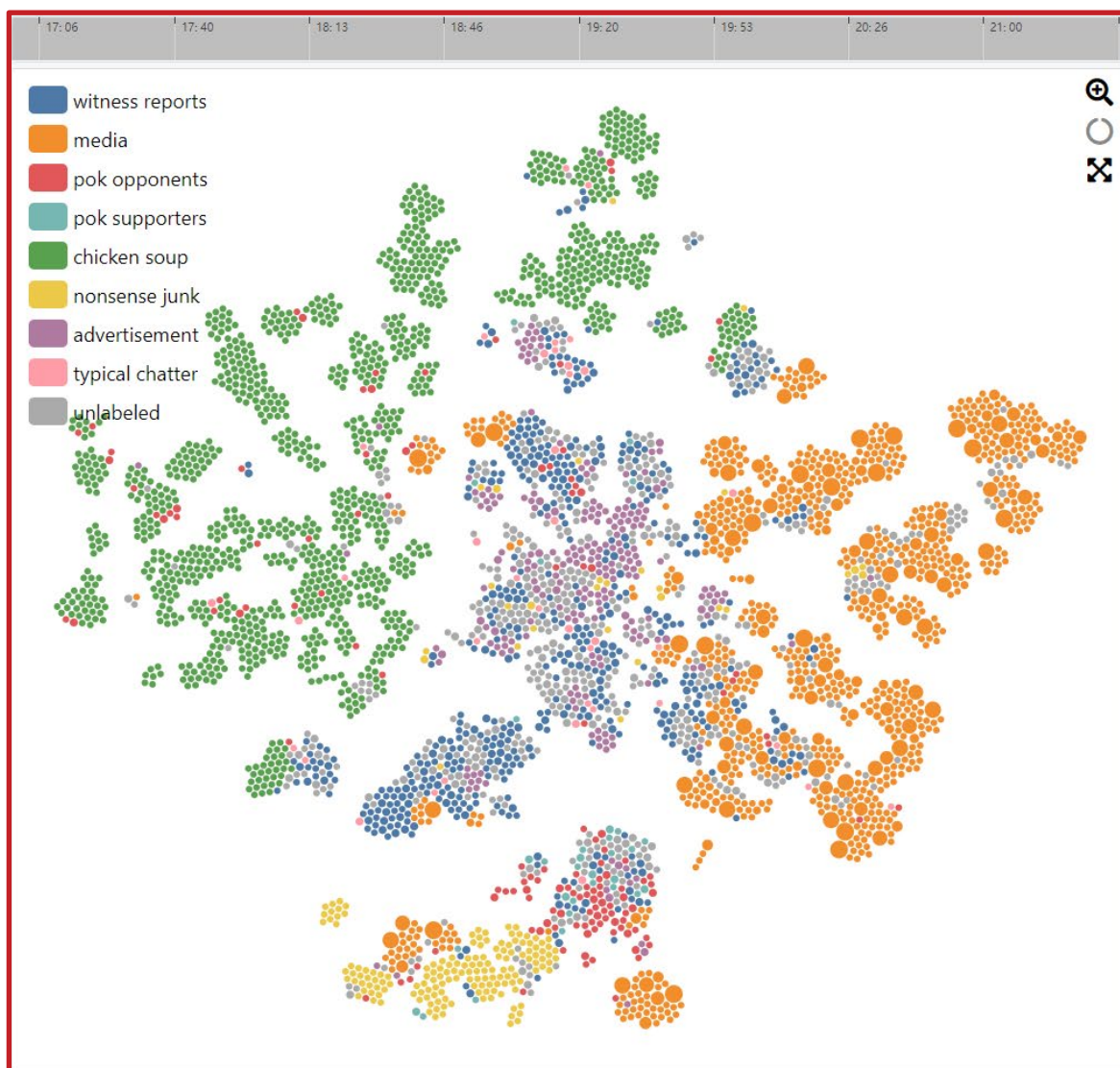


# Three steps to create the message map

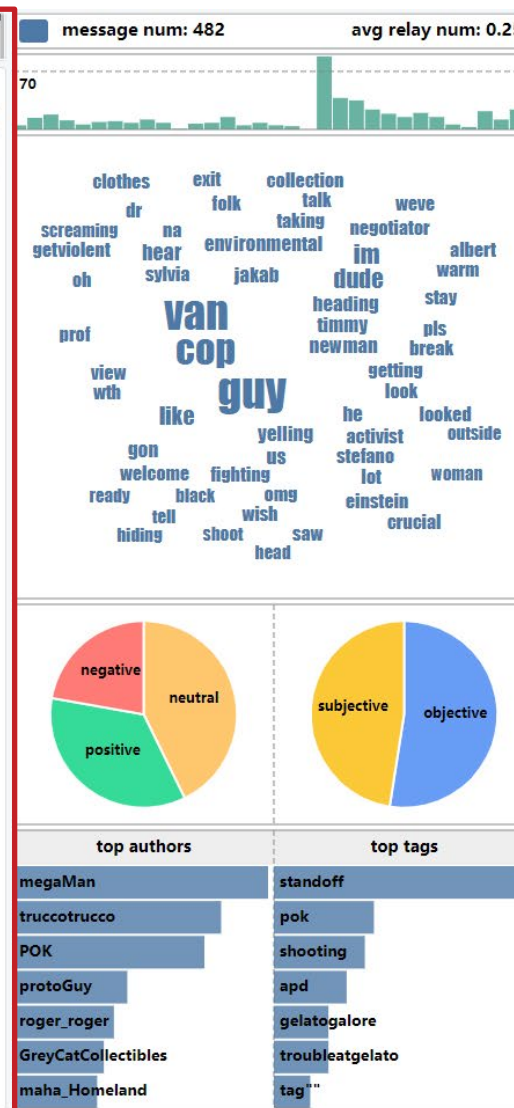


# Second system

message map



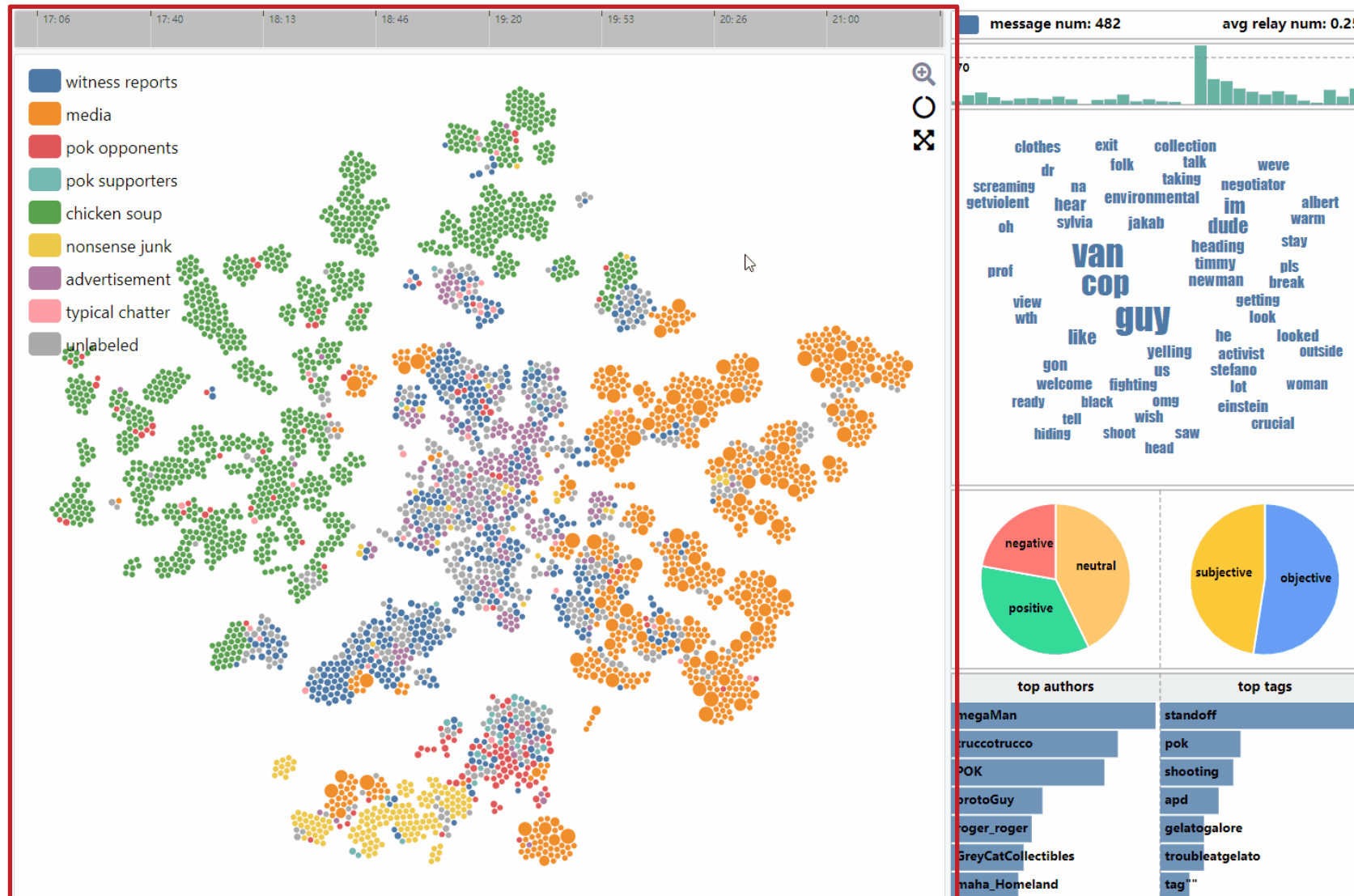
feature checker



# Interactive lens

message map

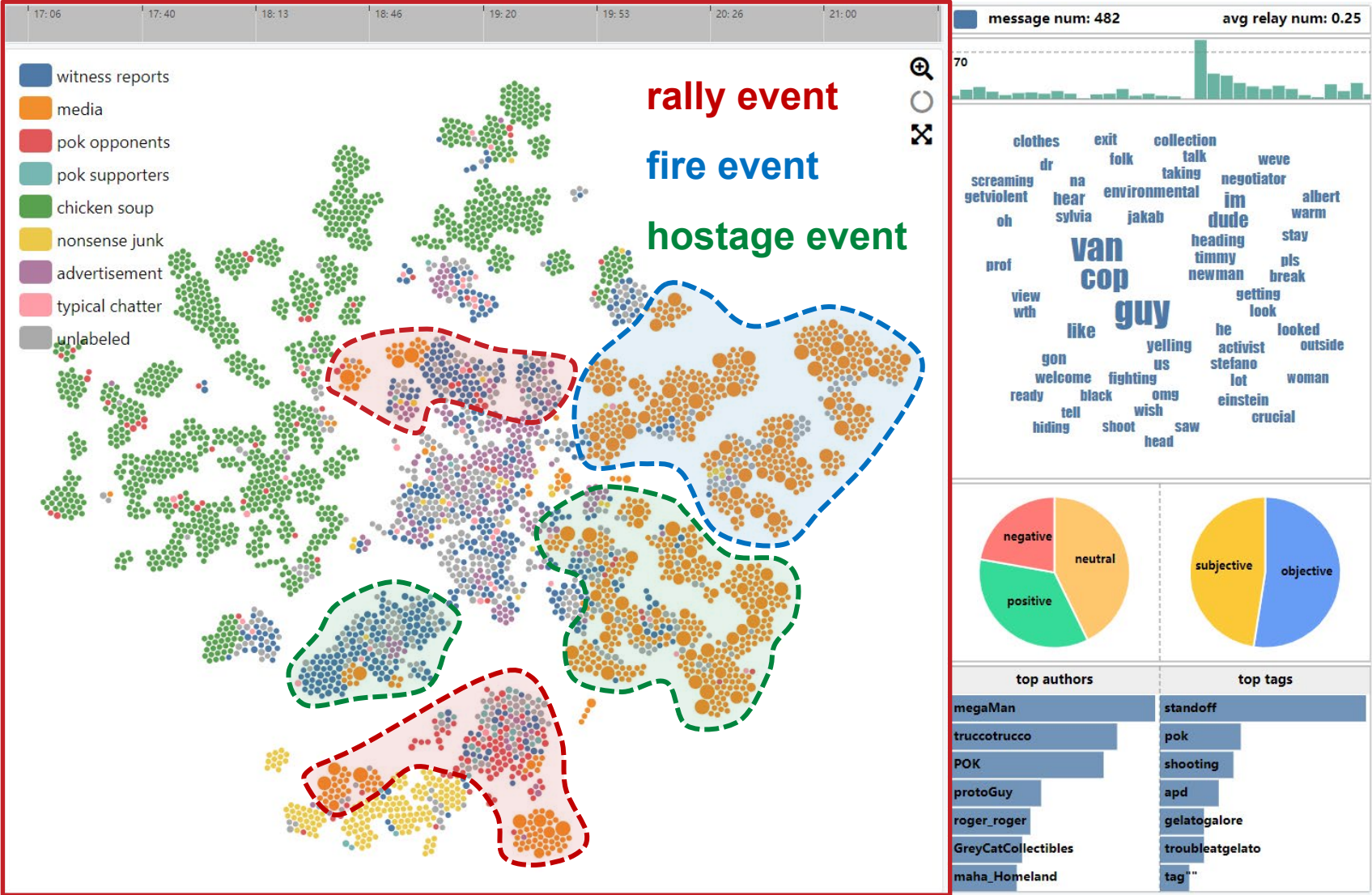
feature checker





# Find messages of the three main events

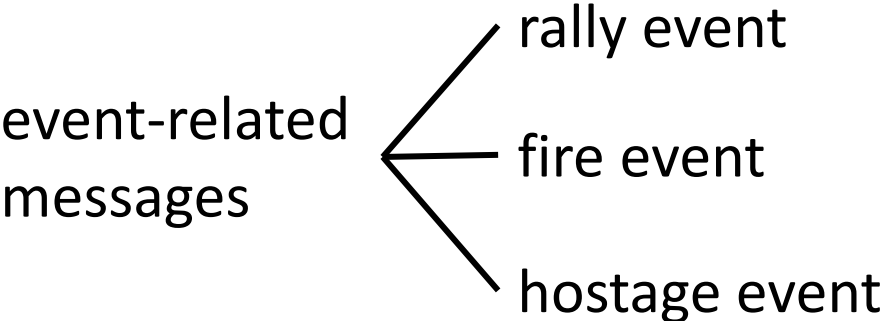
## message map



1. label the circled messages as training data

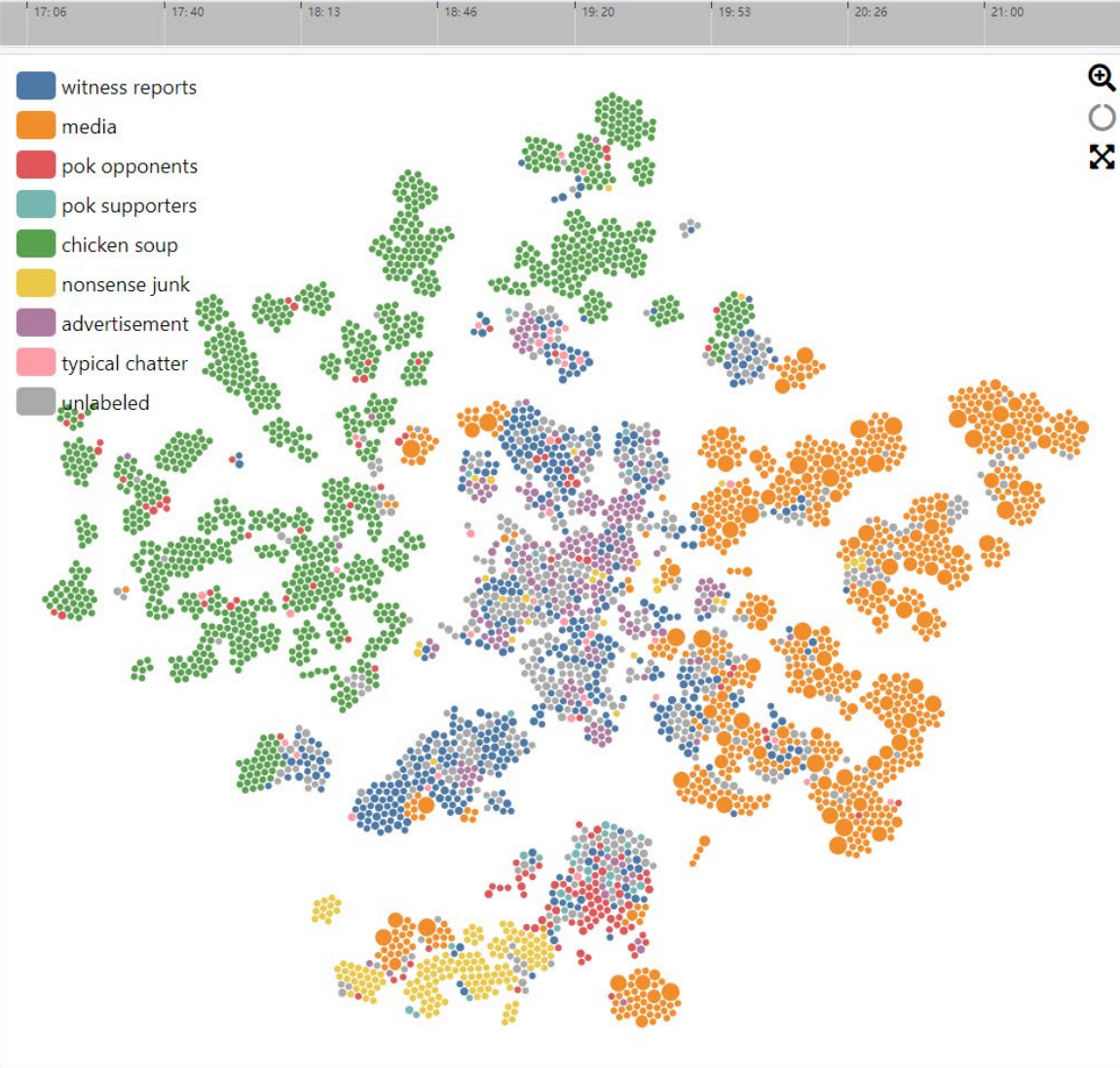
2. train a multi-class classifier

3. apply the classifier to the event-related messages

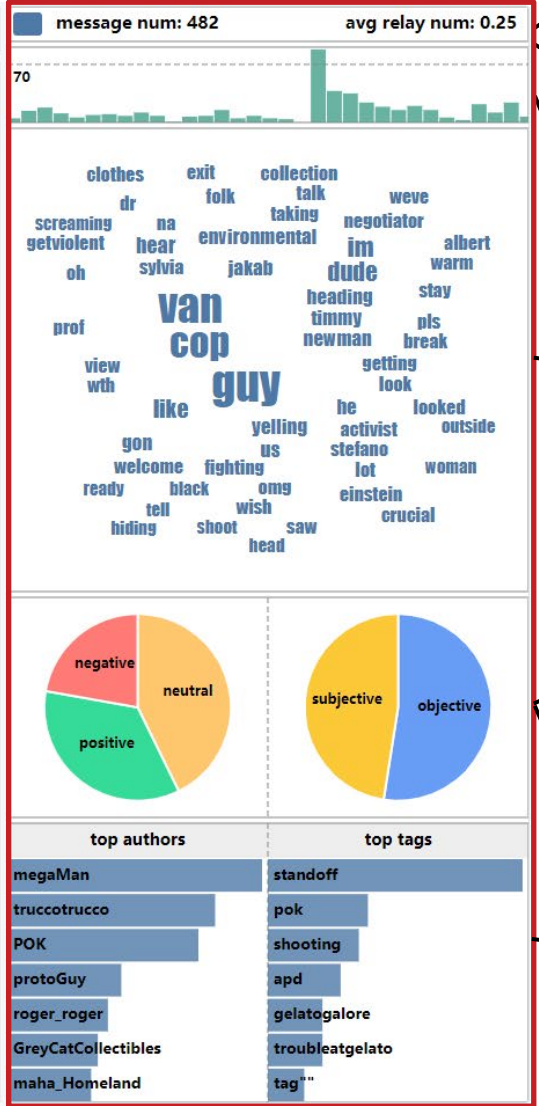


# Features of a class

message map



feature checker



number of messages

forwarded times pre message

distribution of messages over time

feature words

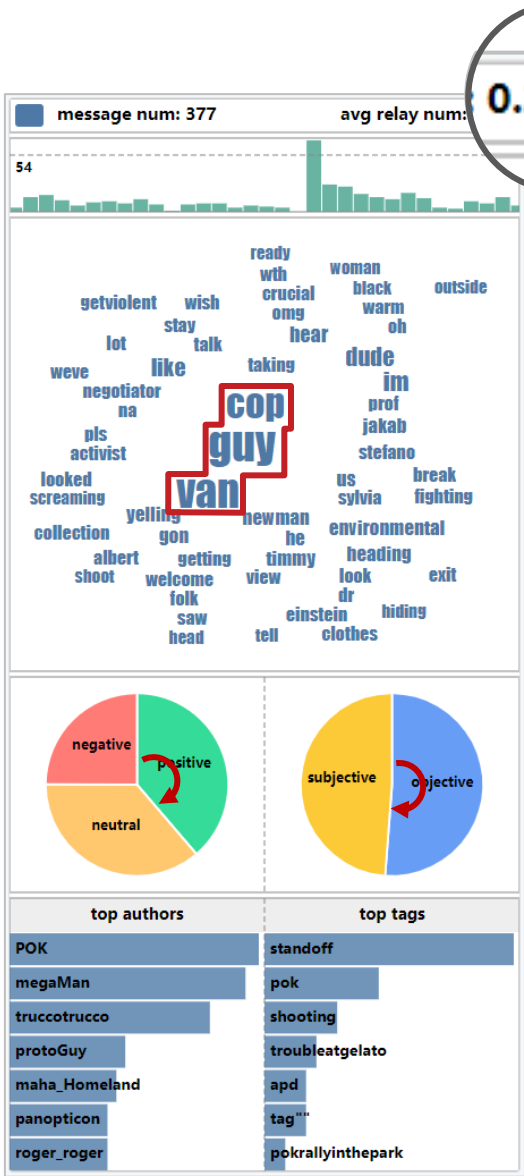
sentiment distribution  
(positive VS neutral VS negative)

subjectivity  
(subjective VS objective)

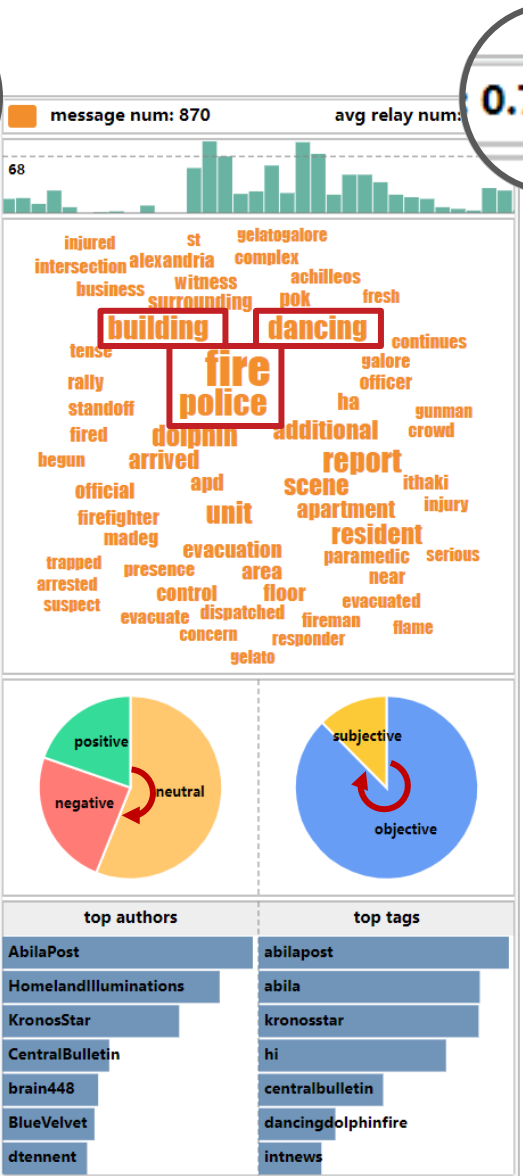
top authors and tags involved

# Differences between witness reports and media

witness reports



forwarded times pre message



media

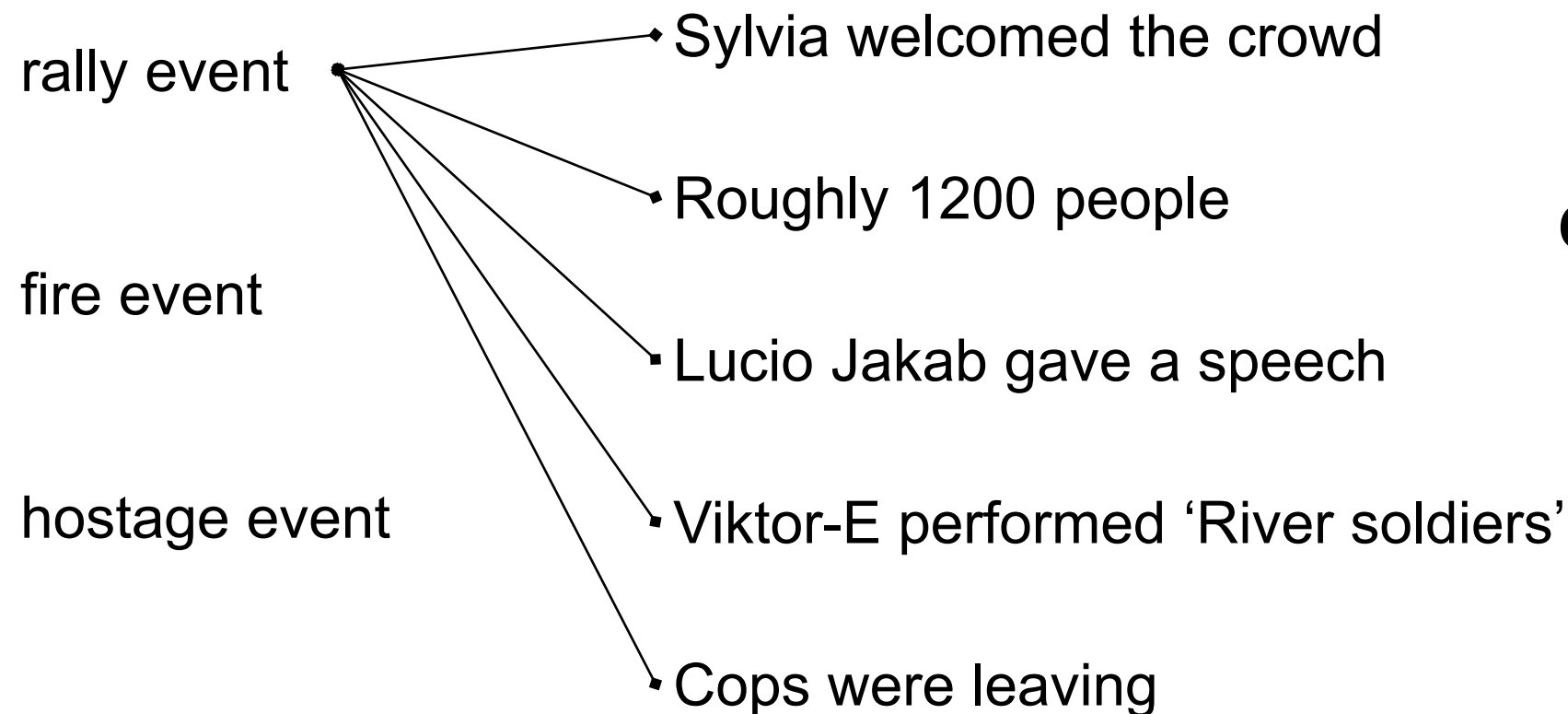
Q2: analyze the evolution of the level of the risk to the public and consider the potential consequences of the situation and the number of people affected



# Unit events

## Main events

## Example of unit events



desired unit events:

- clear semantics
- appropriate granularity

extract by algorithm **FAIL**

human

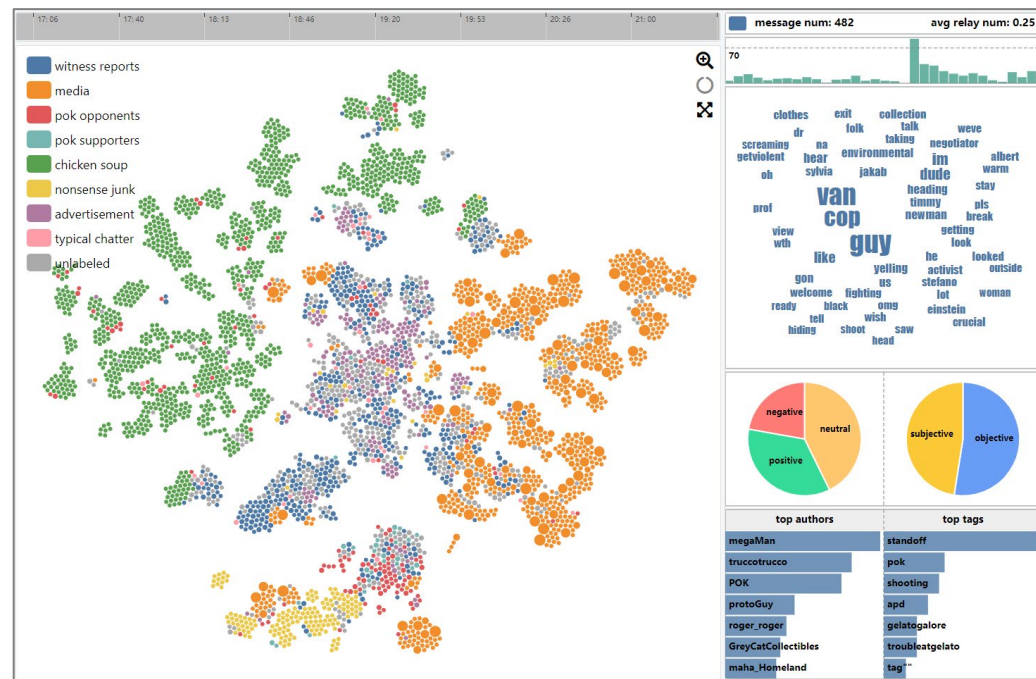
- reasoning
- assessment





# Third system

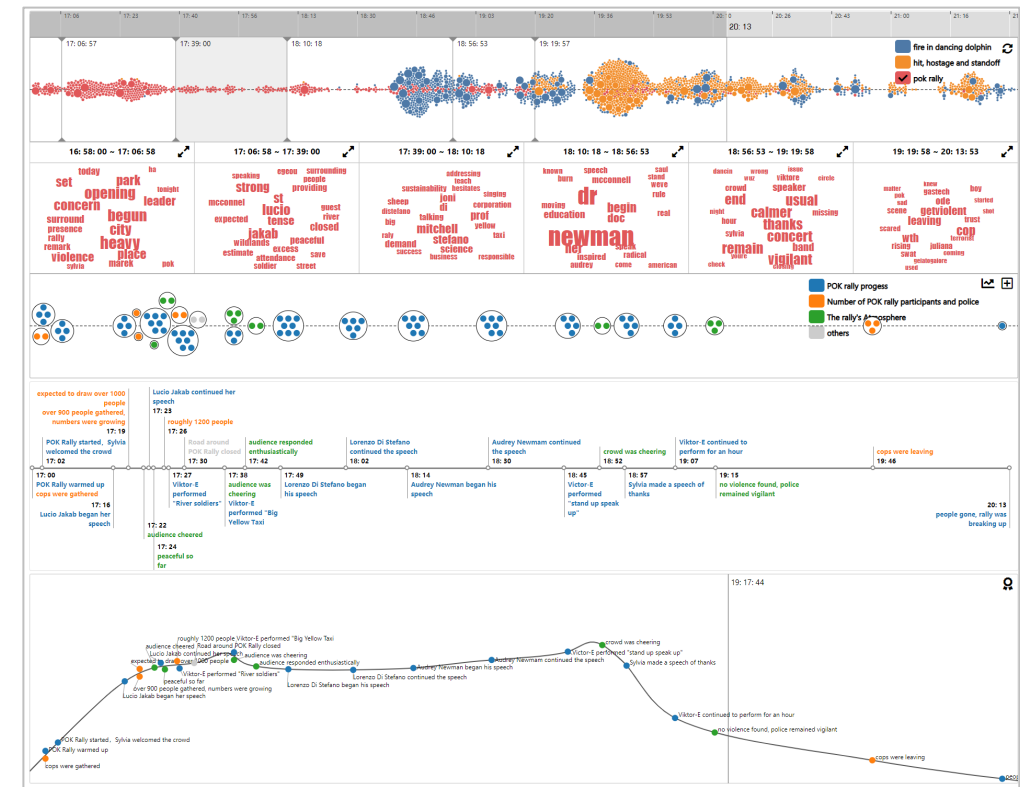
## the second system



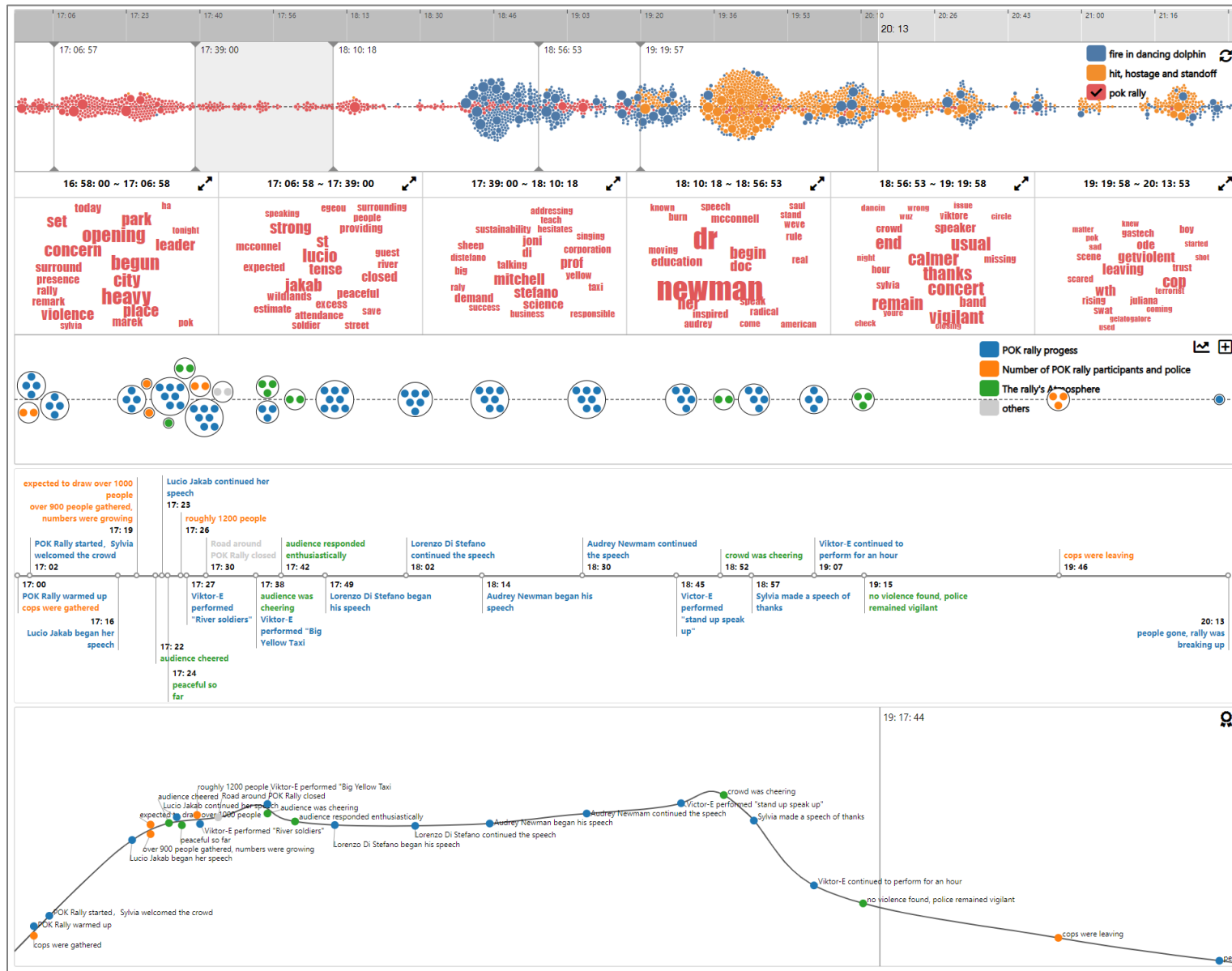
messages of  
the three  
main events



## the third system



# Third system



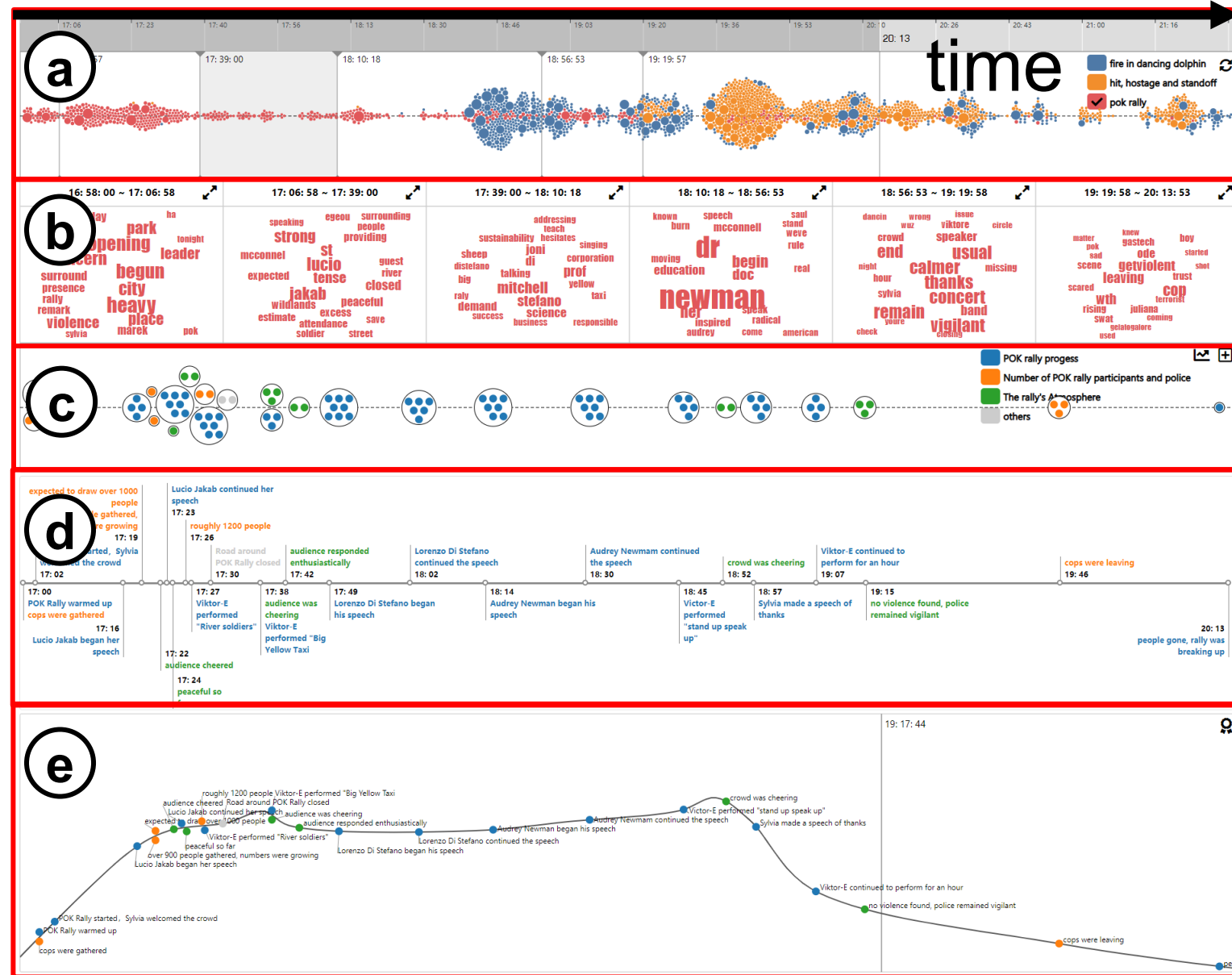
## System properties:

- top-down
- human-led
- narrative visualization creation system

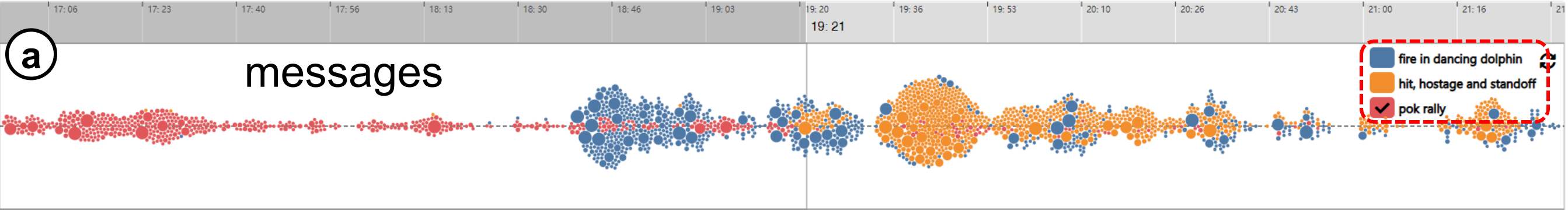
## Steps:

- create unit events
- assess the level of risk

# Third system



# Migration of the risk

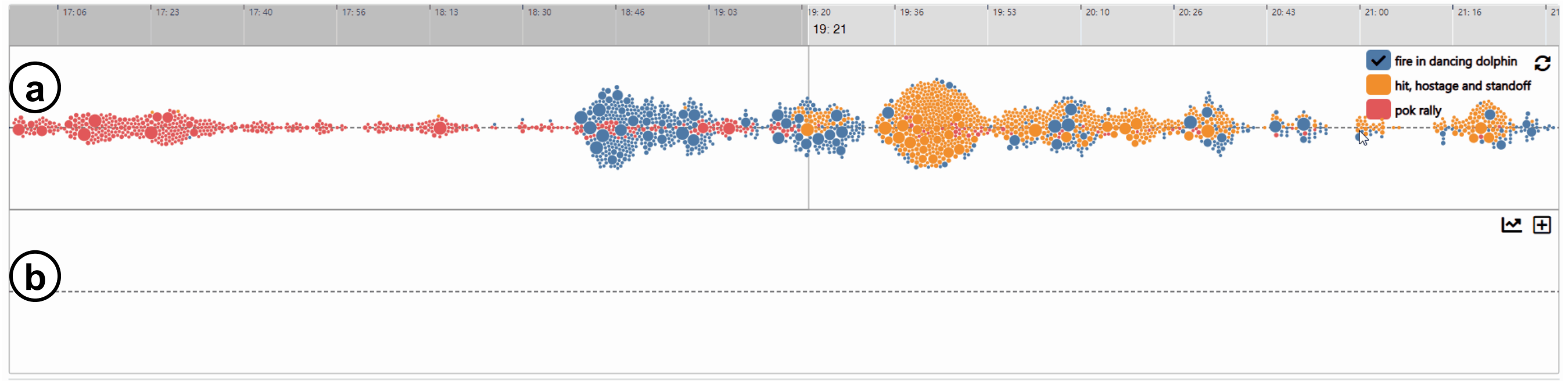


17:00 ← the risk was only from the rally event → 18:39

18:39 ← the risk was concentrated on the fire event → 19:32

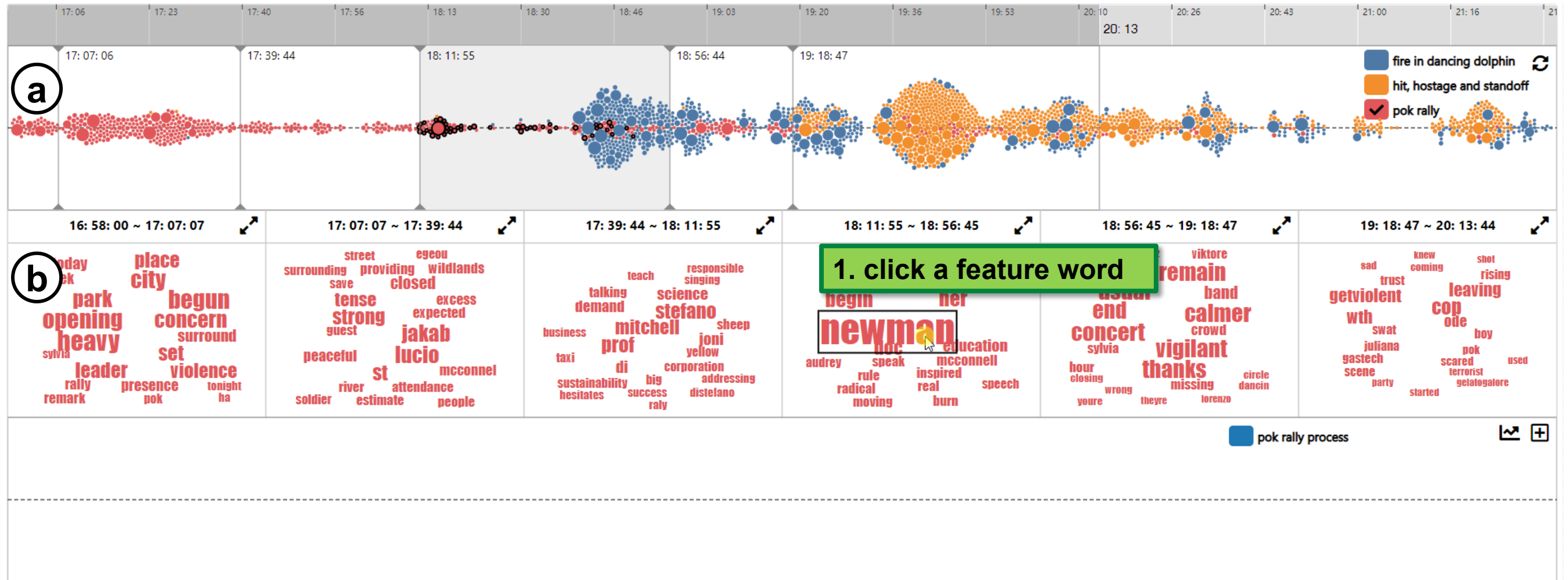
the risk coexisted in the fire event and the hostage event | 19:32 → 21:31

# Create small multiples



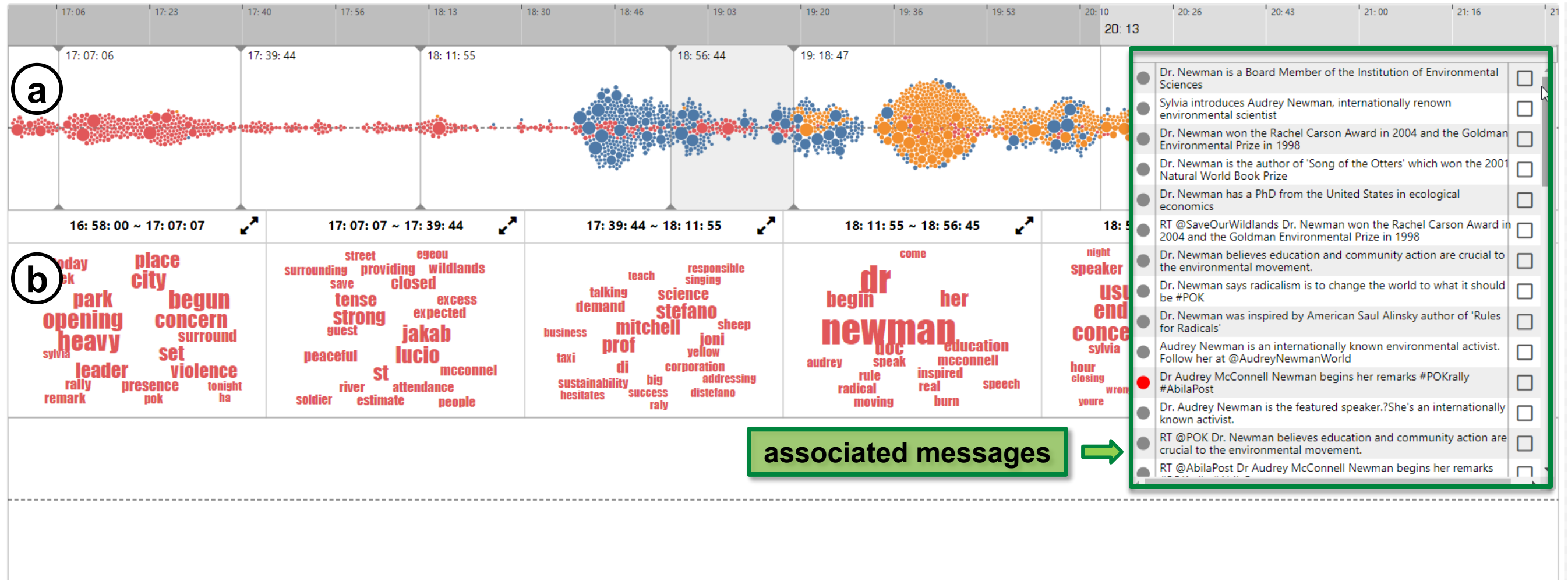
small multiples: feature words

# Steps to create a unit event

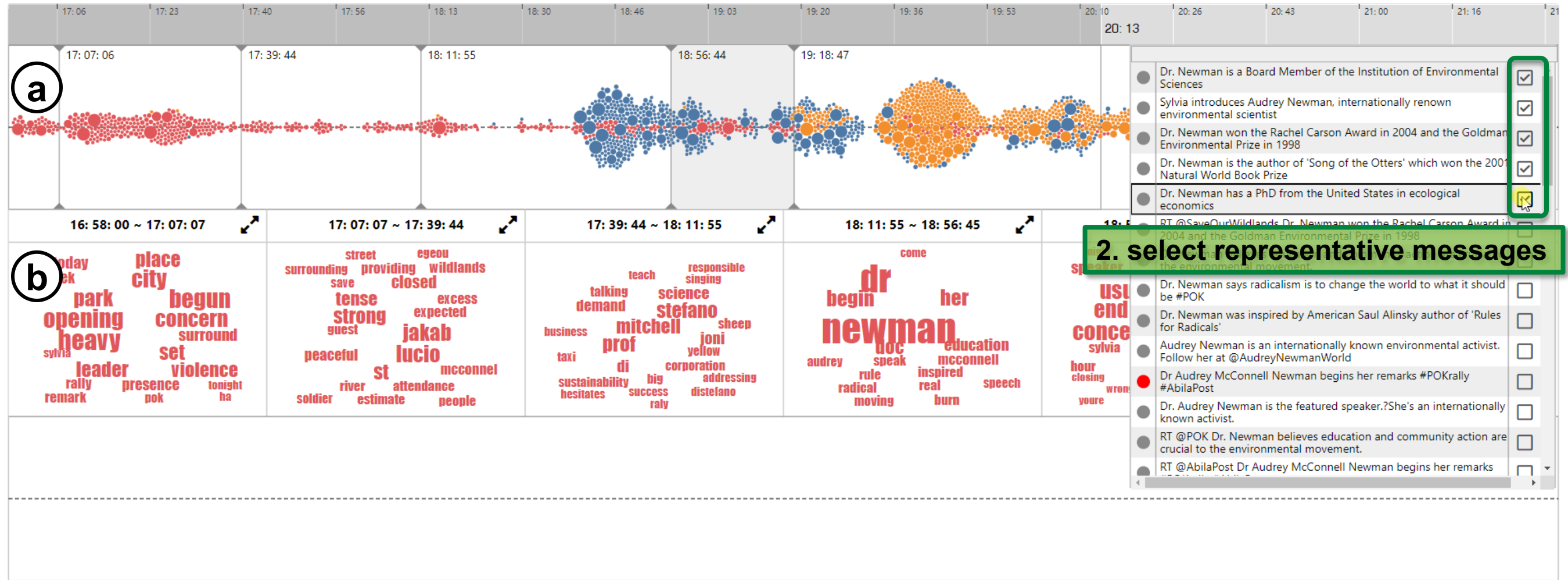




# Steps to create a unit event



# Steps to create a unit event

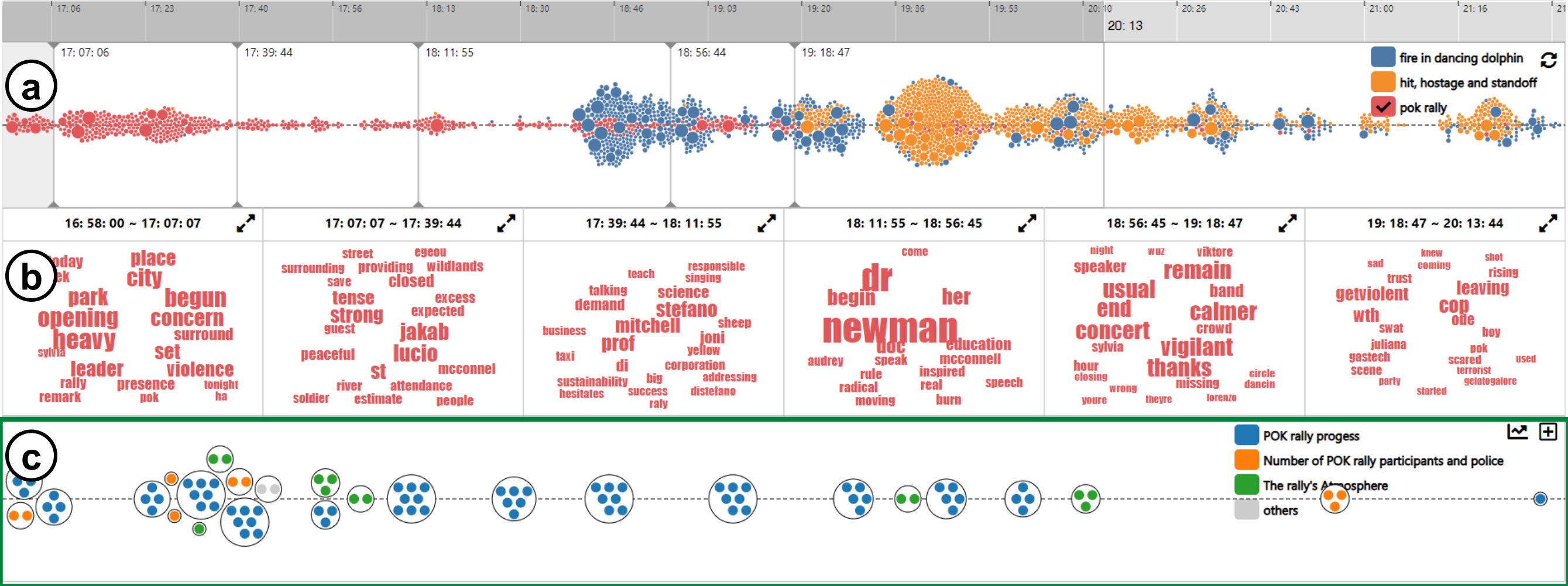




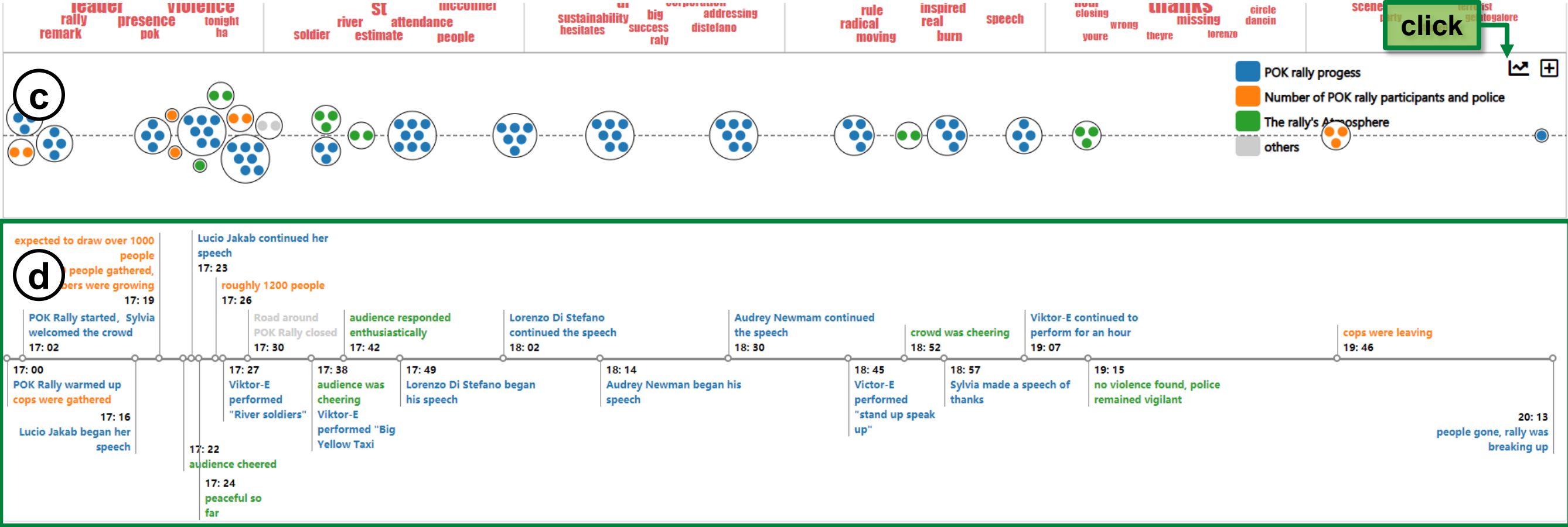
# Steps to create a unit event

The screenshot displays a software interface for creating a unit event. At the top, a timeline shows time intervals from 17:06 to 21:16. Below the timeline, a horizontal bar (a) contains a series of colored dots (red, blue, orange) representing data points over time. Below this, a grid of word clouds (b) is shown for different time intervals, with words like 'park', 'opening', 'heavy', 'leader', 'place', 'city', 'begun', 'concern', 'surround', 'set', 'violence', 'rally', 'presence', 'puk', 'tonight', 'ha', 'soldier', 'river', 'estimate', 'people', 'peaceful', 'st', 'lucio', 'mcconnel', 'surrounding', 'providing', 'wildlands', 'closed', 'tense', 'strong', 'expected', 'excess', 'guest', 'business', 'taxi', 'sustainability', 'big', 'success', 'di', 'corporation', 'addressing', 'distefano', 'joni', 'sheep', 'yellow', 'mitchell', 'prof', 'stefano', 'science', 'responsible', 'singing', 'teach', 'demand', 'talking', 'begin', 'her', 'education', 'mcconnell', 'audrey', 'rule', 'speak', 'inspired', 'real', 'speech', 'night', 'speaker', 'ust', 'end', 'conce', 'sylvia', 'hour', 'closing', 'wron', 'youre'. A text input field (c) contains the text ". Newman gave a speech" with "Close" and "Submit" buttons. A green box highlights the text "3. name the unit event". A red box highlights a cluster of blue dots, with an arrow pointing to it and the text "a representative message". Below this, an arrow points to a small icon of a cluster of blue dots with the text "the created unit event". On the right side, a list of messages is displayed, each with a checkbox. The messages include: "Dr. Newman is a Board Member of the Institution of Environmental Sciences", "Sylvia introduces Audrey Newman, internationally renown environmental scientist", "Dr. Newman won the Rachel Carson Award in 2004 and the Goldman Environmental Prize in 1998", "Dr. Newman is the author of 'Song of the Otters' which won the 2001 Natural World Book Prize", "Dr. Newman has a PhD from the United States in ecological economics", "RT @SaveOurWildlands Dr. Newman won the Rachel Carson Award in 2004 and the Goldman Environmental Prize in 1998", "Dr. Newman believes education and community action are crucial to the environmental movement.", "Dr. Newman says radicalism is to change the world to what it should be #POK", "Dr. Newman was inspired by American Saul Alinsky author of 'Rules for Radicals'", "Audrey Newman is an internationally known environmental activist. Follow her at @AudreyNewmanWorld", "Dr Audrey McConnell Newman begins her remarks #POKrally #AbilaPost", "Dr. Audrey Newman is the featured speaker.?She's an internationally known activist.", "RT @POK Dr. Newman believes education and community action are crucial to the environmental movement.", "RT @AbilaPost Dr Audrey McConnell Newman begins her remarks".

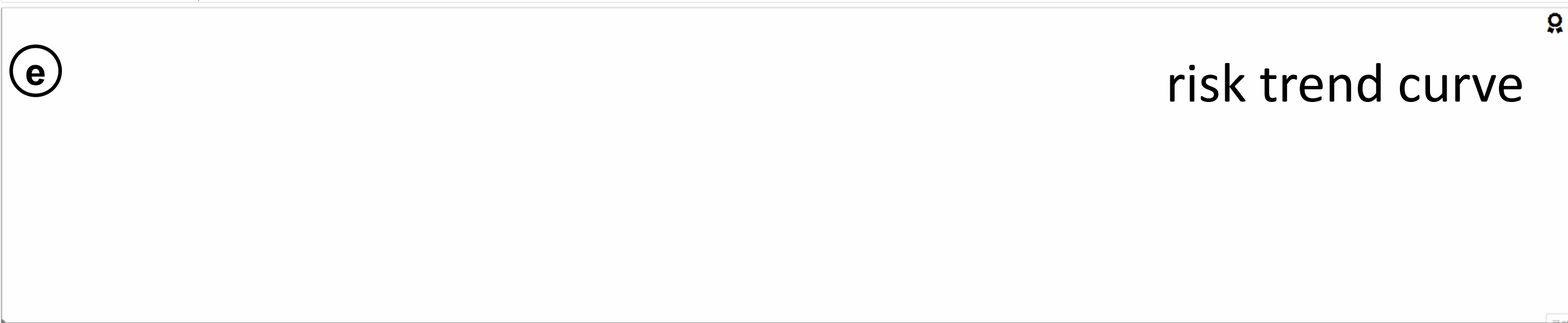
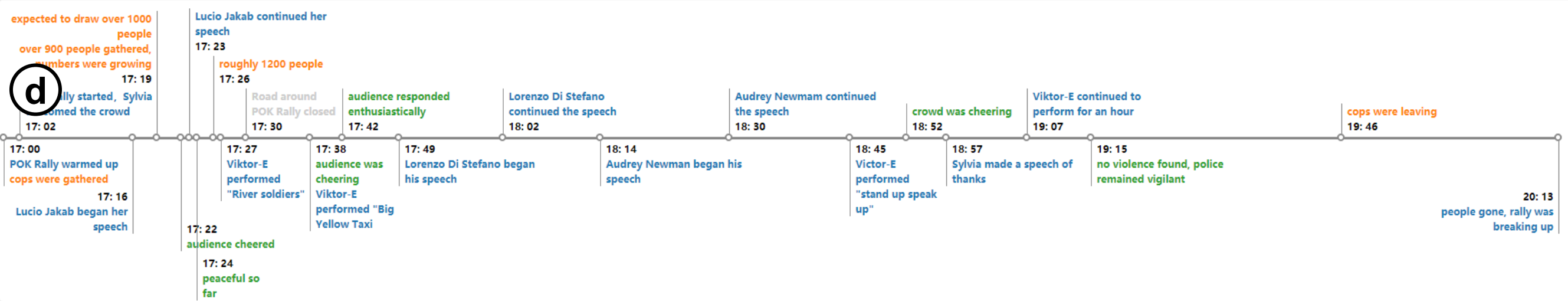
# Narrative visualization: a timeline consisting of unit events



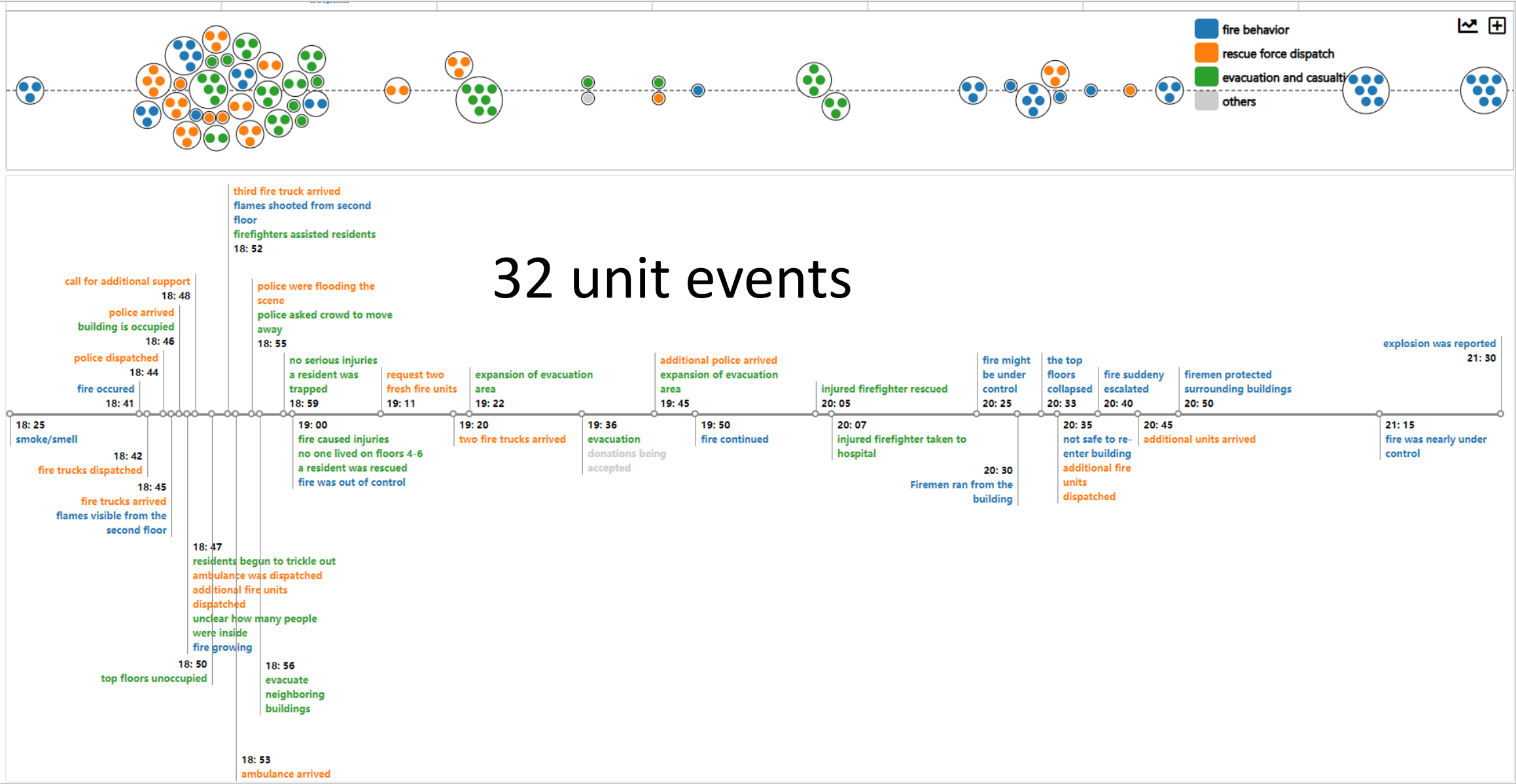
# Narrative visualization: a timeline with explicit event names



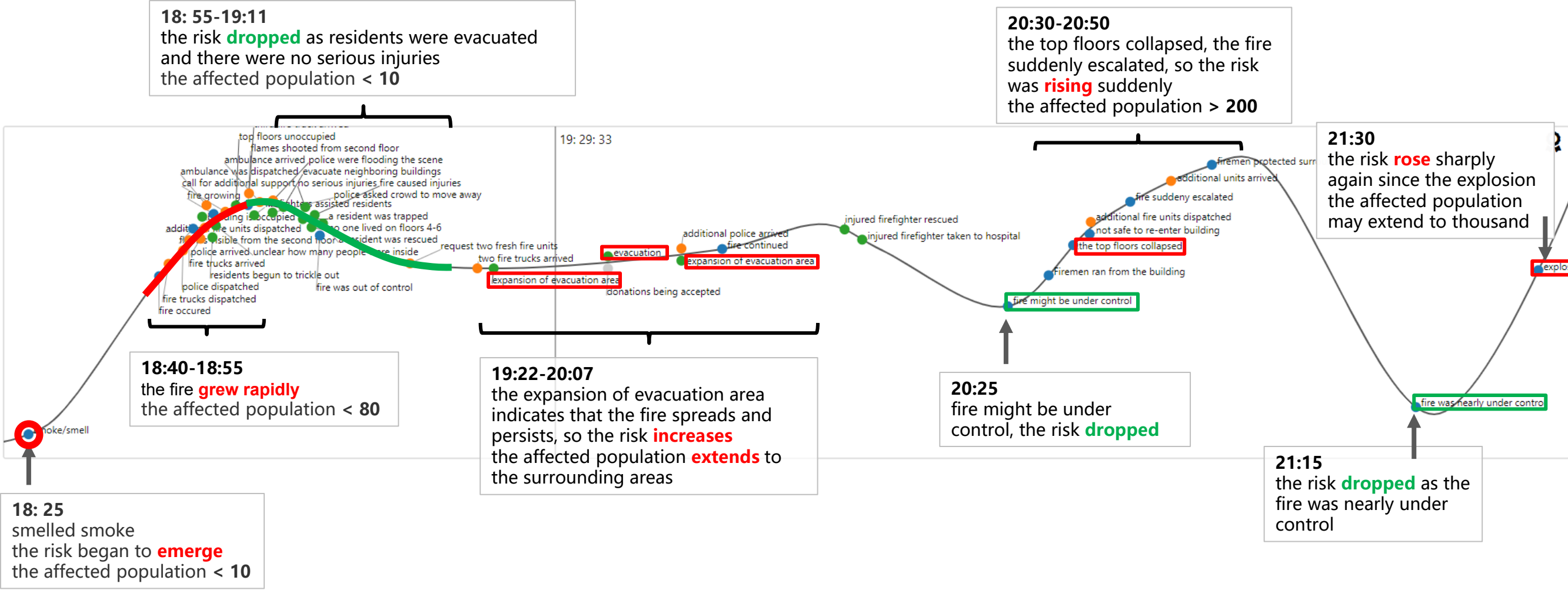
# Narrative visualization: risk trend curve



# Risk assessment of the fire event



# Risk assessment of the fire event



Q3: make decisions about where to dispatch a team of first responders in real-time analysis and retrospective analysis

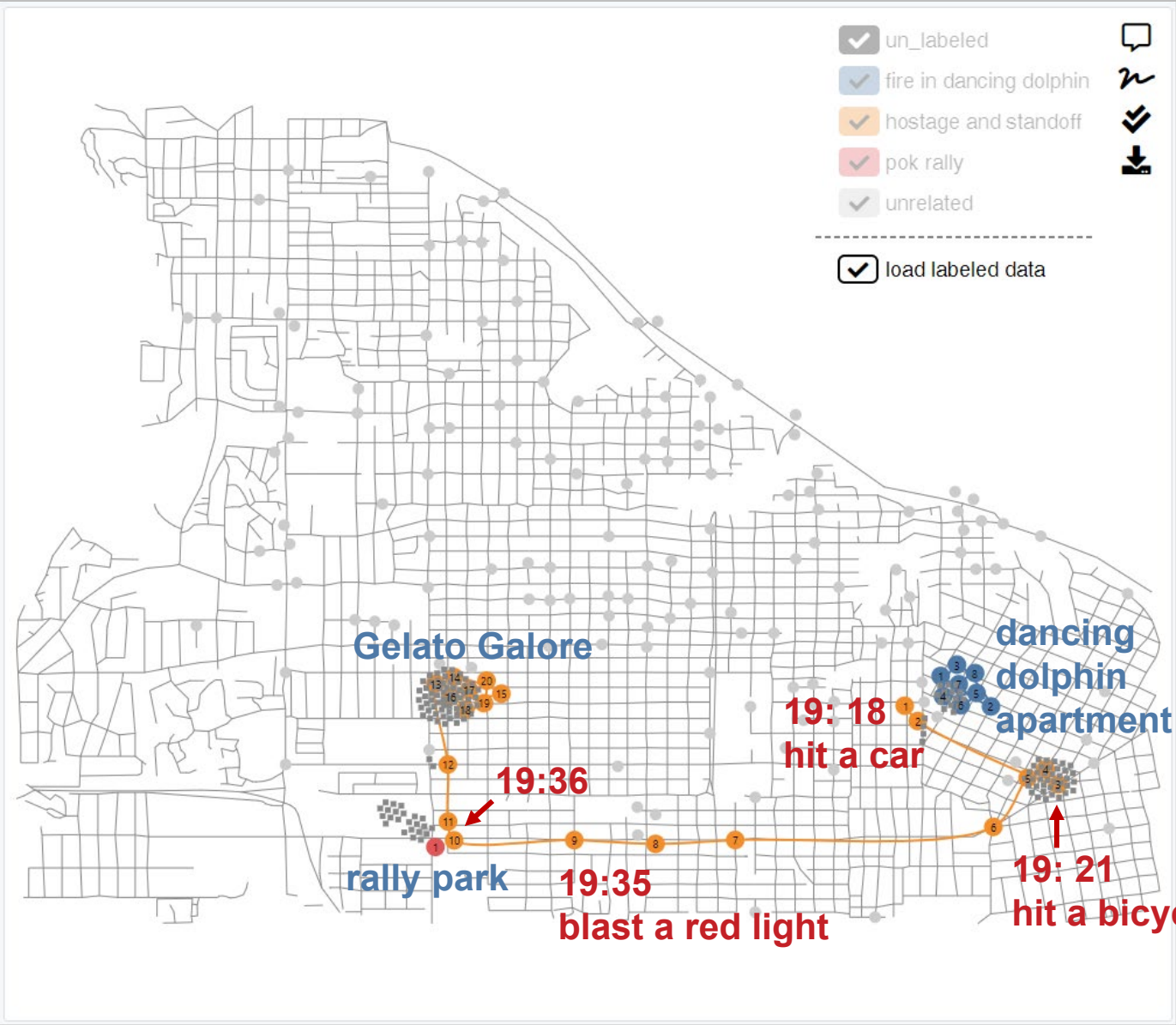
# From a retrospective perspective: fired apartment

Principle: send the team to the location of the event with the most serious consequences

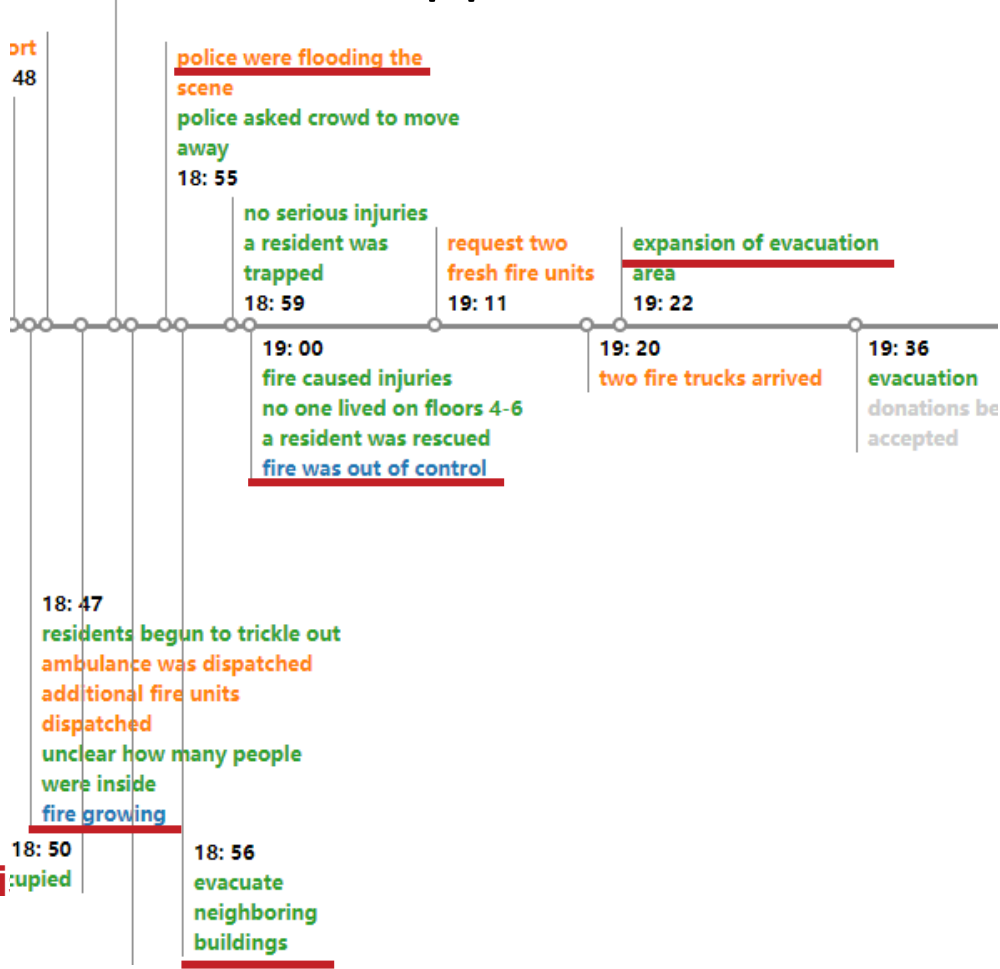
- real-time analysis: the potential and presumable consequences
- retrospective analysis: the actual consequences



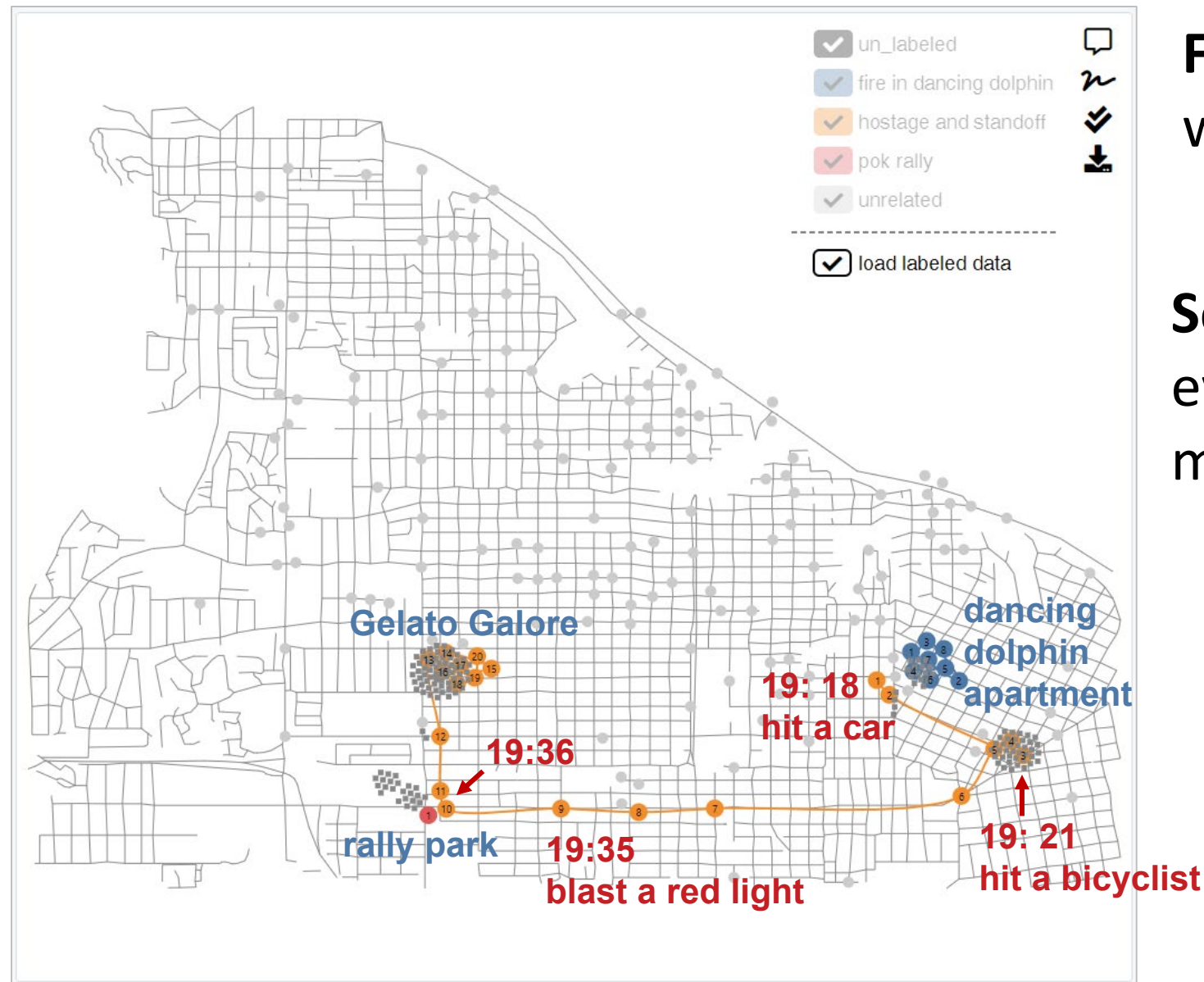
# From a retrospective perspective: fired apartment



First, we suspect that the hostage event would not happen without the fire event.



# From a retrospective perspective: fired apartment



**First**, we suspect that the hostage event would not happen without the fire event.

**Second**, compared to the other two events, the fire event actually has the most serious consequences.

# From a real-time perspective: rally park

- **First**, the potential for mass violence at the rally is high.
- **Second**, in the event of mass violence, the cost would be enormous.
- **Third**, the rally precedes the other two events and its risk is predictable.

Q4: compare the solution of 2014  
and the current solution

# Differences between two solutions

	solution in 2014	solution in 2021
<b>Scenario</b>	real-time, streaming data	retrospective, top-down
<b>Who plays the lead role</b>	system	analyst
<b>What the system provides</b>	visualizations in fixed template	creation tools for a narrative visualization
<b>What does the analyst need to do</b>	read and understand the given visualizations	structuring data as evidence, recording and analyzing evidence, and formulating conclusions as visualizations
<b>How to record conclusions</b>	remain in mind or be recorded in words	be drawn as a narrative visualization
<b>Interactions</b>	limited	rich

# Conclusion

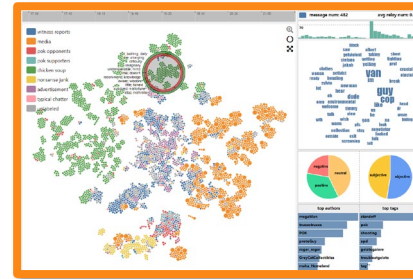
#1



classified  
messages



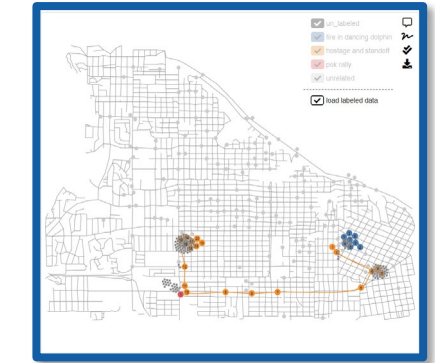
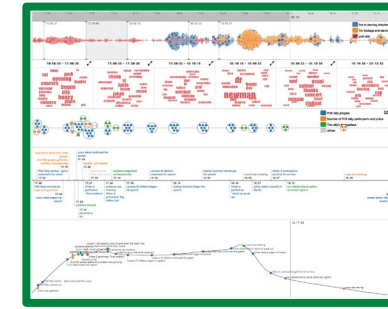
#2



messages of  
three main  
events

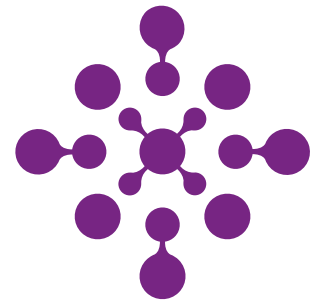


#3



- Providing analysts with a flexible exploration environment at the beginning often yields twice the results with half the effort.
- The idea of interactive creation of narrative visualization is worth trying, especially for the problem that the required abstract semantics cannot be derived from algorithms.





VIS 2021

---

Thanks

