

RAI SOCIAL: UN'ANALISI

SETTIMANALE DEI

TWEET



Alex Ceccotti – 790497
Pietro De Simoni – 790638
Gianmarco Stucchi – 778772



OBIETTIVO
E TASK DA RISOLVERE



- X Programmi TV Rai
- X Twitter
- X Gestione dati streaming
- X Integrazione con dati delle pagine ufficiali

È necessario rispondere a domande business oriented



- ✓ Architettura che permetta la gestione del flusso di dati
- ✓ Analisi real-time
- ✓ Analisi descrittiva batch
- ✓ Scalabilità

È importante fornire delle visualizzazioni comprensibili dai manager

2.

PROGRAMMAZIONE TV

Scraping

<		Venerdì 20	Giovedì 21	Venerdì 22	Oggi 23	Domani 24	Lunedì 25	Martedì 26	Mercoledì 27	Giovedì 28	>	
<		12:00	12:30	13:00	13:30	>						
In Diretta												
Rai 1		Dreams Road Story - SudAfrica 11:35 - 40 min		Linea Verde va in città - In Puglia 12:20 - 70 min		TELEGIORNALE 13:30 - 30 min		Linea Blu - Rimini 14:00 - 90 min				
Rai 2		No Tomorrow S1E5 - Nessun rim... 10:05 - 40 min		No Tomorrow S1E6 - I debiti van... 10:45 - 40 min		La nave dei sogni - Tanzania 11:25 - 95 min						
i.eventSpan:nth-of-type(n+25) div.eventWrapper						ettimanale Estate 12:00 - 18 min						
Rai 3						Correva l'anno 12:55 - 65 min						
						TG Regione 14:00 - 18 min						

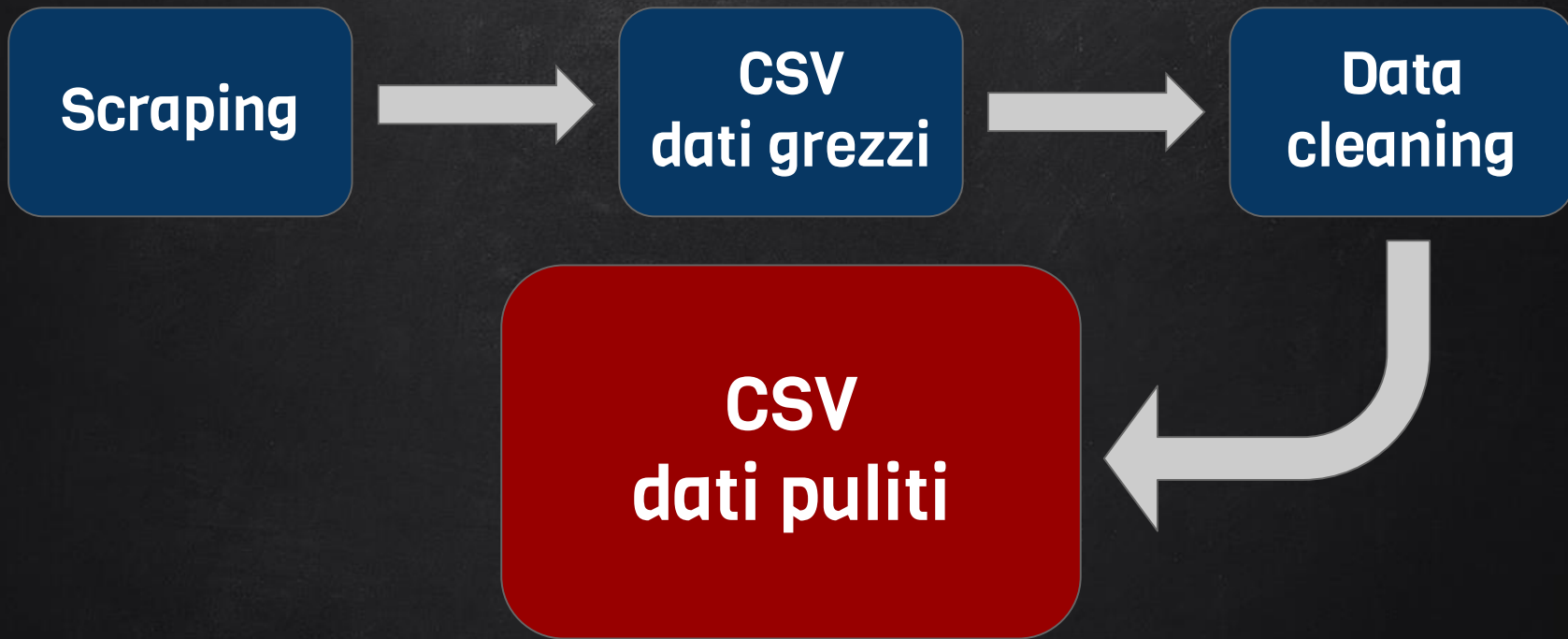
Scraping



```
graph TD; A[Scraping] --> B[CSV dati grezzi]
```

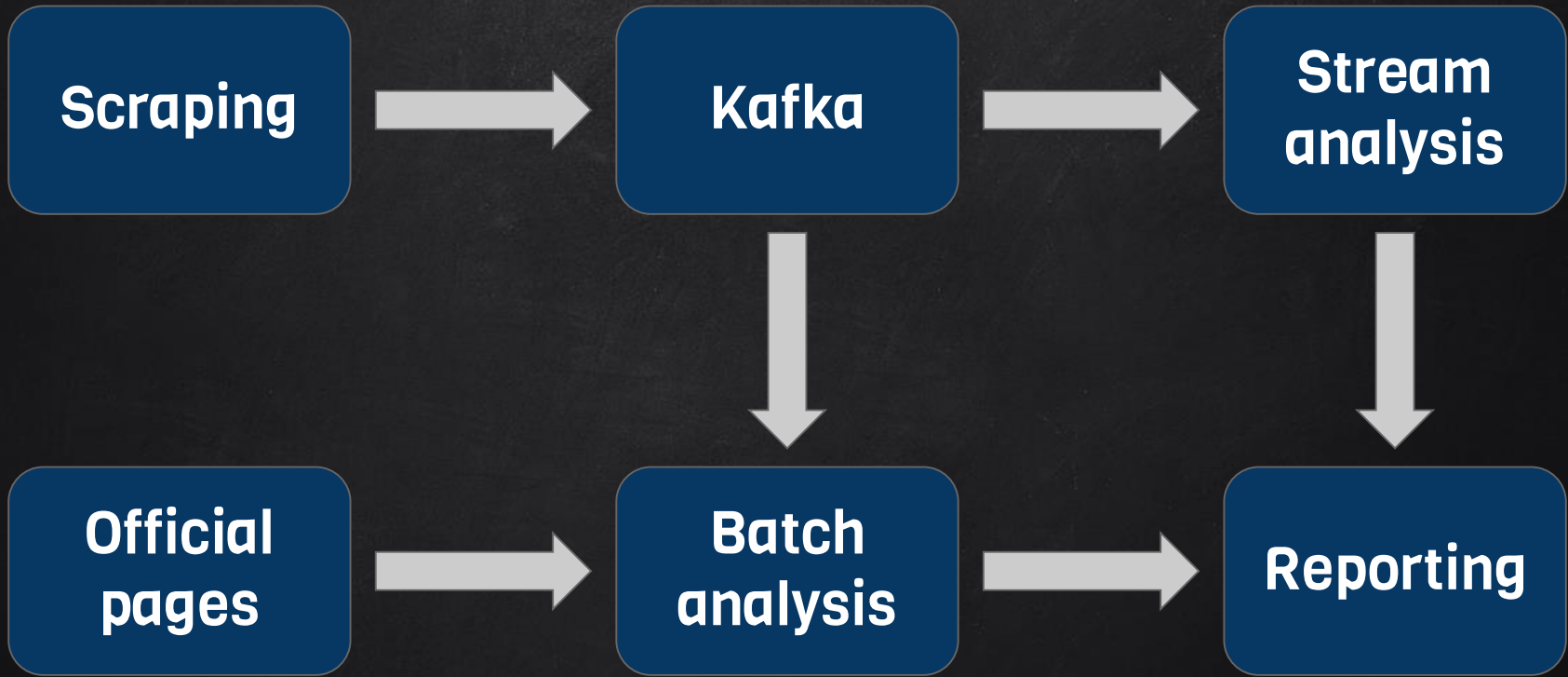
**CSV
dati grezzi**

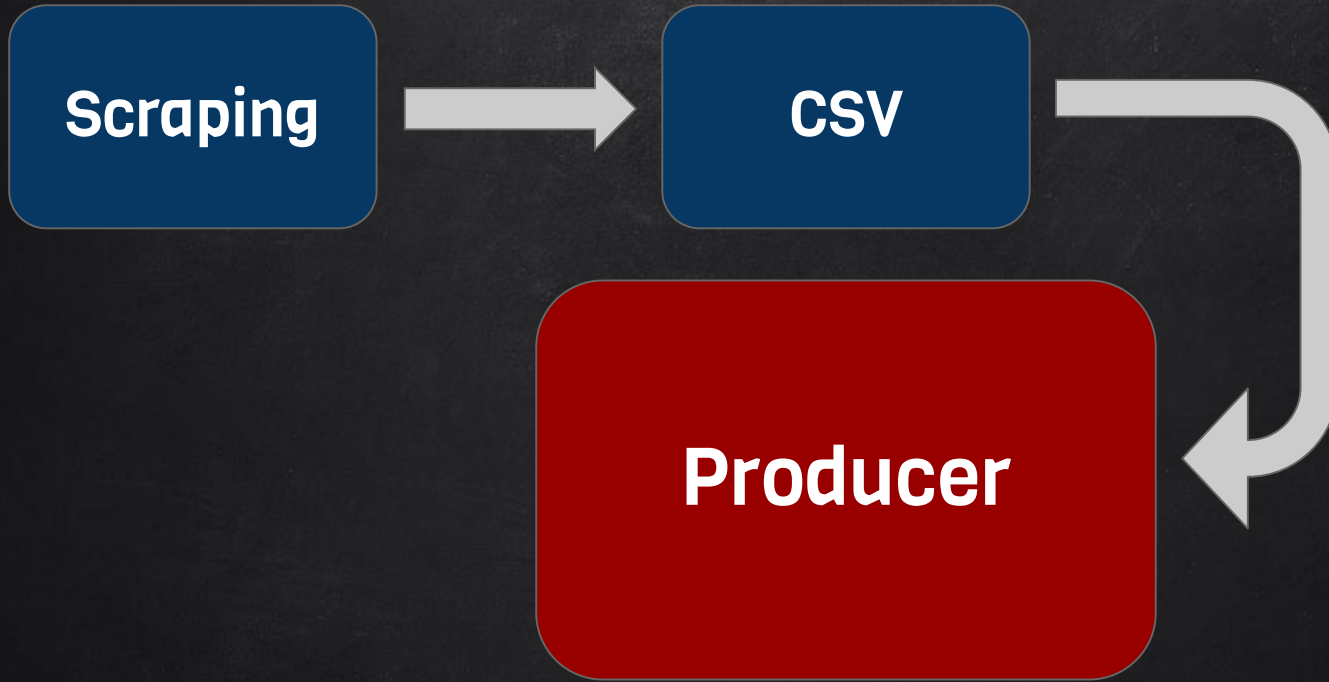


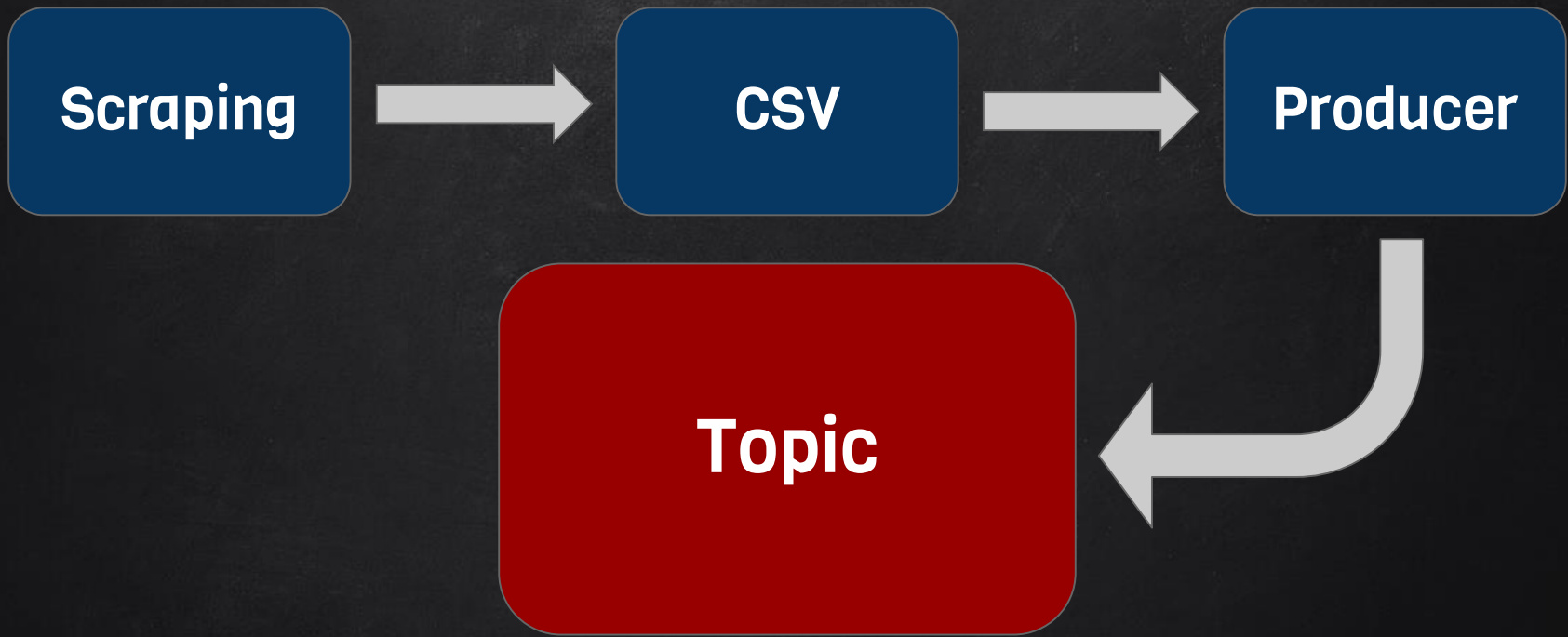


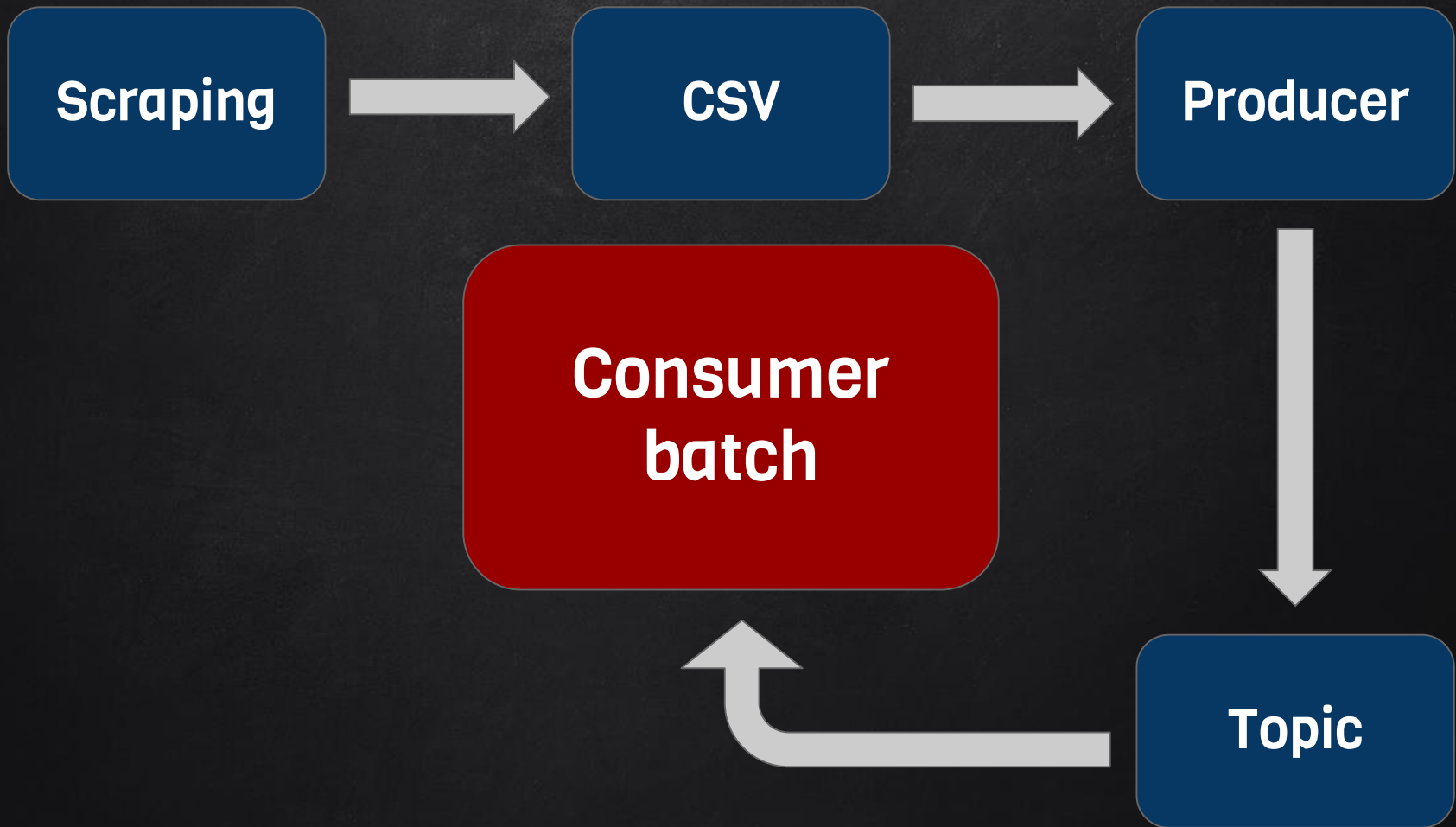


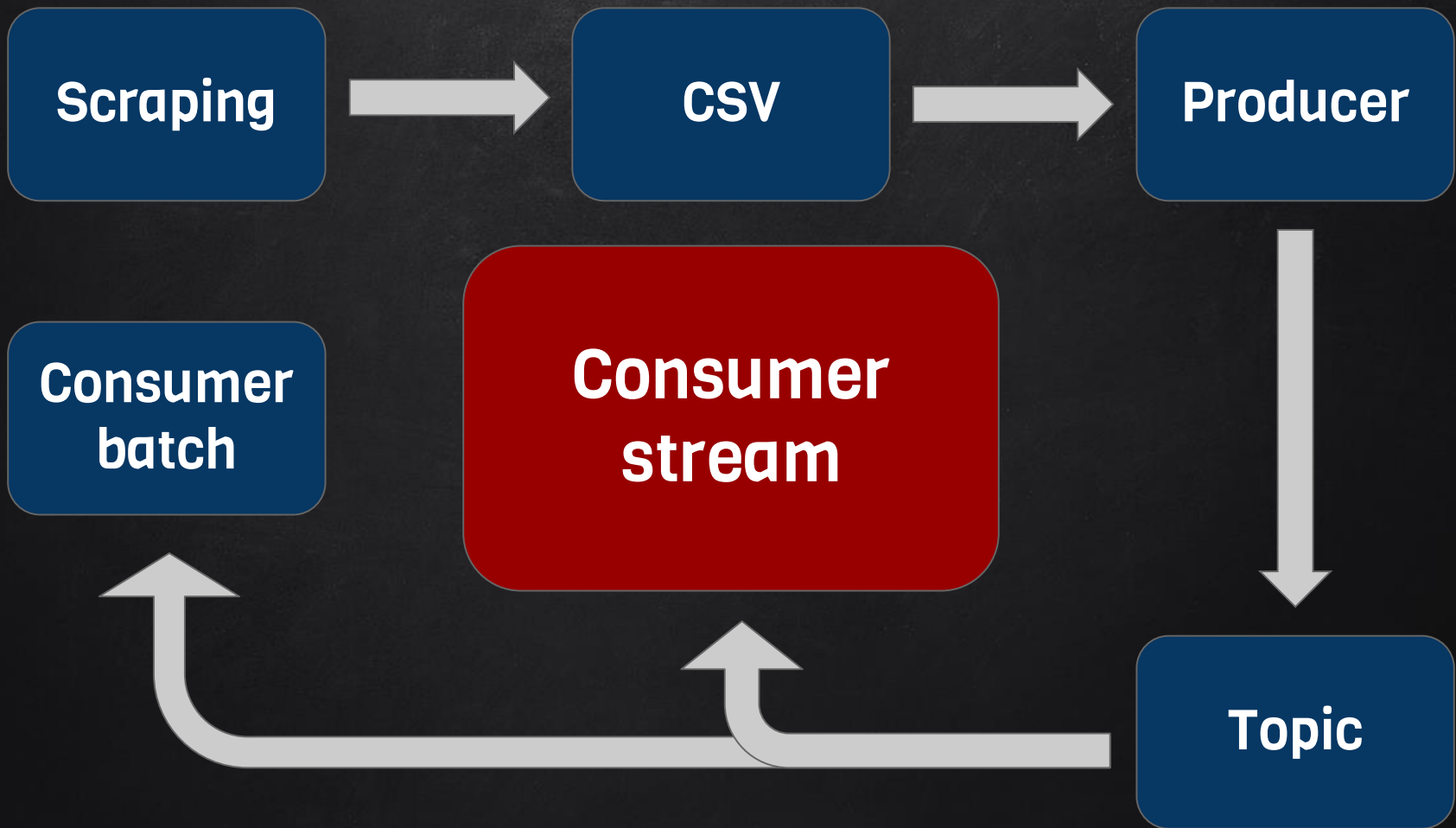
ARCHITETTURA











**Consumer
batch**

**Consumer
stream**

**Consumer
batch**



MongoDB

**Consumer
stream**

**Consumer
batch**



MongoDB

**Consumer
stream**

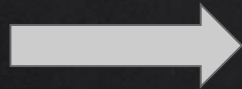
- Id
- Screen name
- Created at
- Text
- Followers
- Friends
- Hashtag
- Nome evento
- Orario inizio
- Orario fine
- Keywords

**Consumer
batch**



MongoDB

**Consumer
stream**



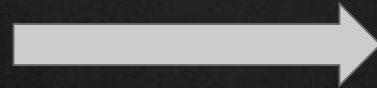
**Hashtag
più usati**

**Consumer
batch**



MongoDB

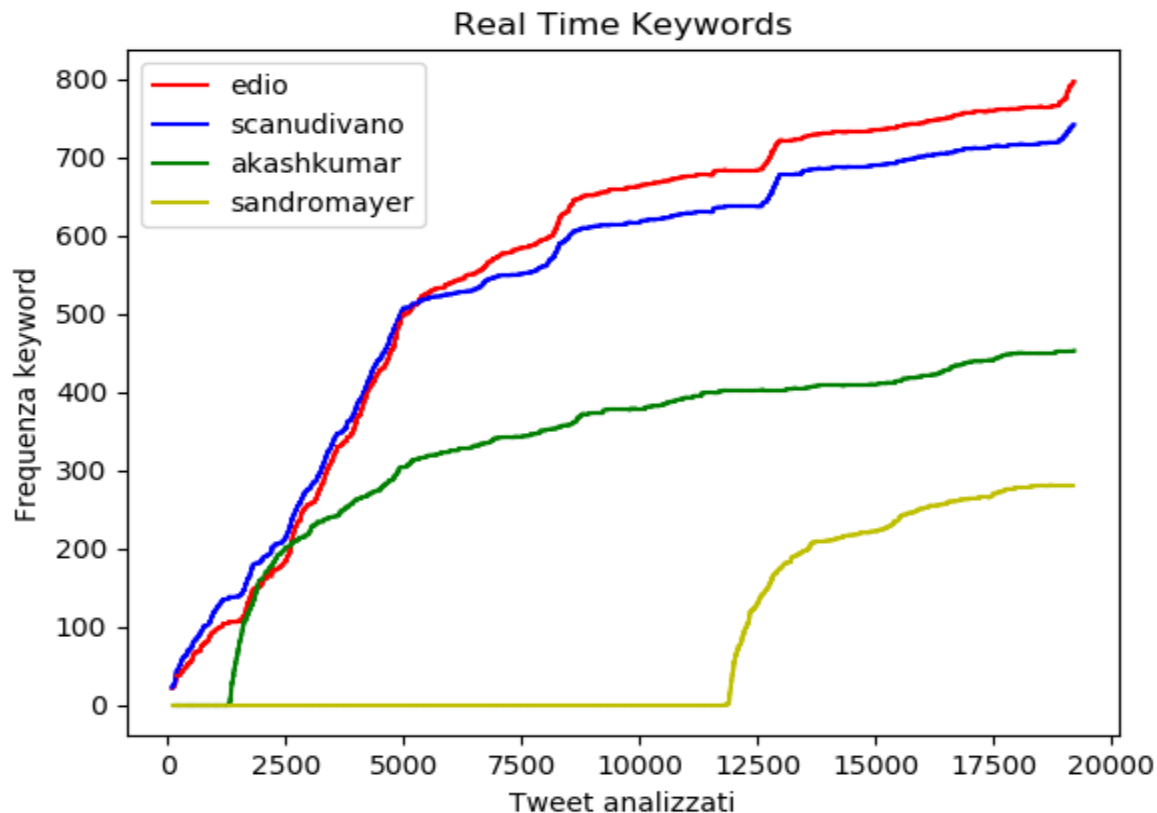
**Consumer
stream**



Hashtag



ANDAMENTO KEYWORDS DURANTE “BALLANDO CON LE STELLE”

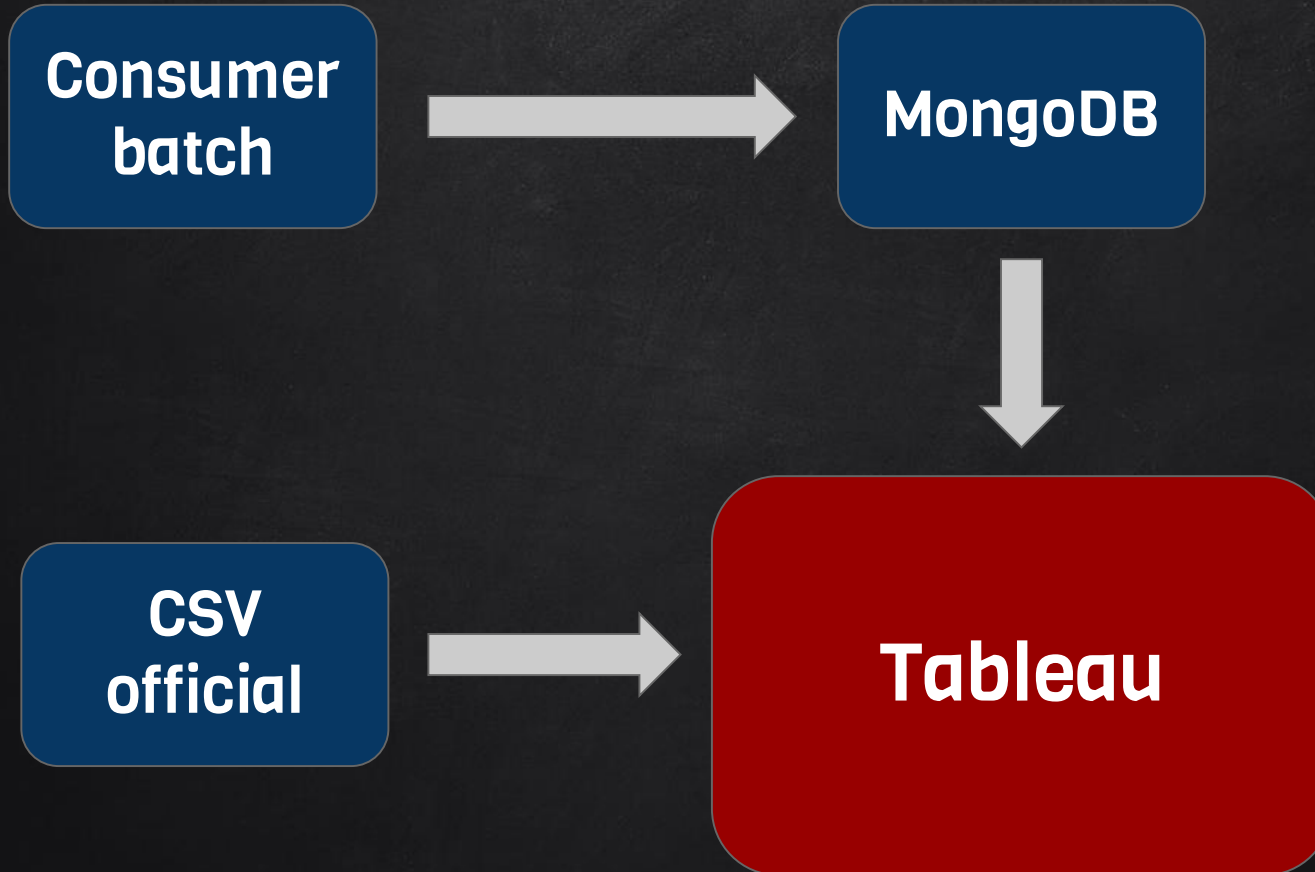


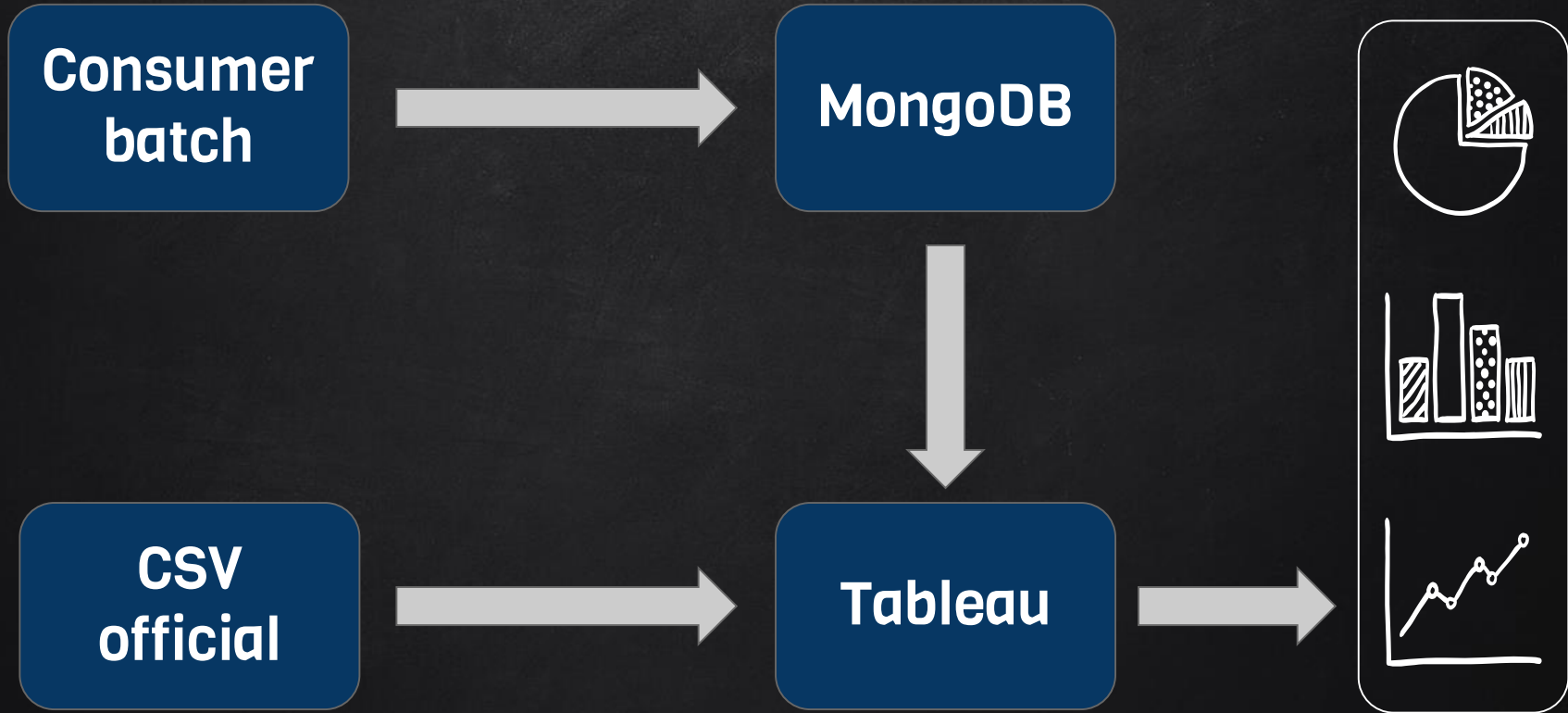
**Consumer
batch**



MongoDB

**CSV
official pages**

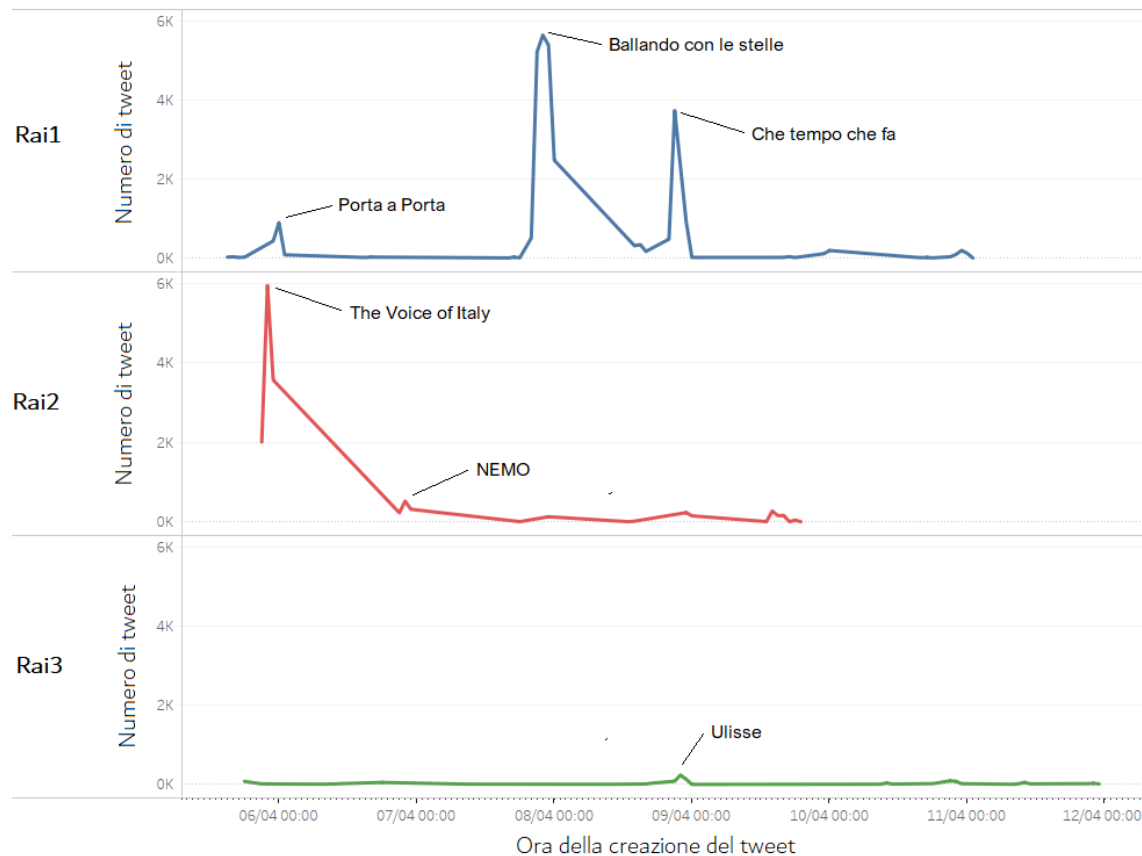




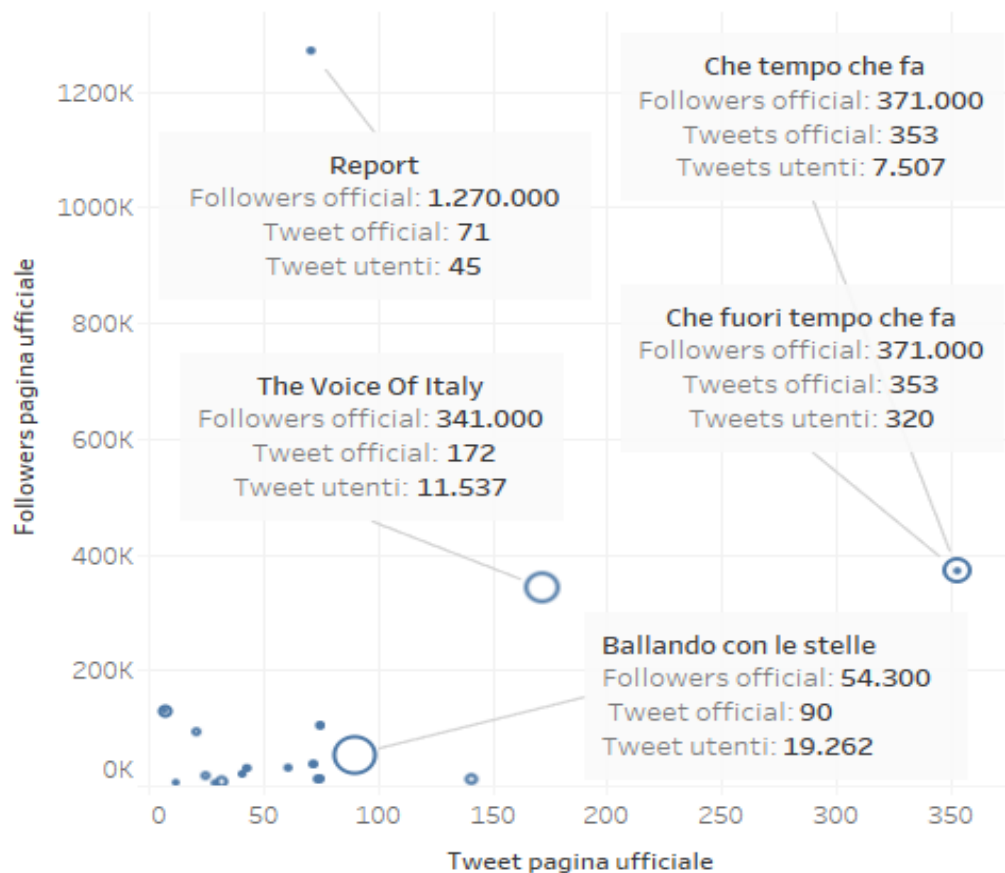


DATA ANALYSIS

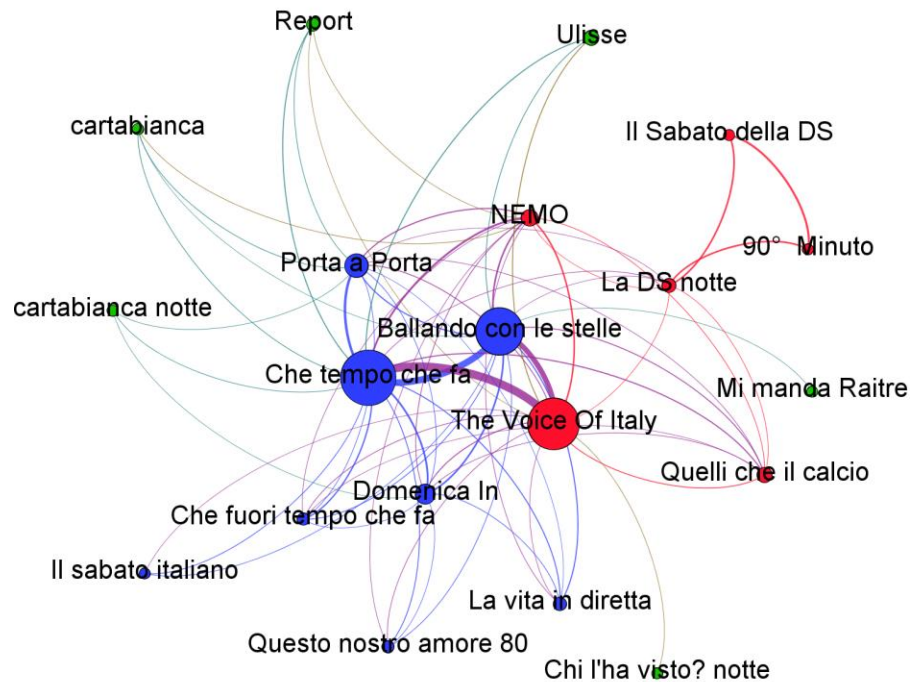
ANDAMENTO ORARIO DEI TWEET PER CANALE

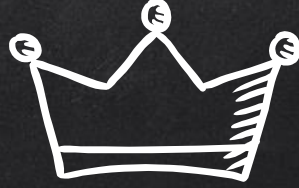


RELAZIONE TRA PAGINE UFFICIALI E TWEET



RELAZIONE TRA GLI EVENTI





GRAZIE PER
L'ATTENZIONE!



Alex Ceccotti – 790497
Pietro De Simoni – 790638
Gianmarco Stucchi – 778772