Giambattista Amati, Simone Angelini, *(Fondazione Ugo Bordoni)*

Anna Caterina Carli, Giuseppe Pierri *(Istituto Superiore delle Comunicazioni e delle Tecnologie dell'Informazione)*

Giorgio Gambosi, Daniele Pasquini, Gianluca Rossi *(Università Tor Vergata Roma)*

Paola Vocca *(Università della Tuscia - VT)*

Analisi temporale degli eventi su Twitter

Twitter: temporal events analysis

Sommario*: Obiettivo di questo lavoro di ricerca è la classificazione degli eventi in base a differenti pattern temporali, corrispondenti al picco del volume dei messaggi scambiati, per comprendere come gli eventi si propagano sui social network Twitter-like. In prima battuta viene fornita una definizione puntuale di "eventi unici", strettamente correlata al concetto di hashtag. Prendendo in considerazione specifici intervalli di tempo, gli hashtag più popolari sono selezionati tramite la tecnica Seasonal Hybrid ESD (S-H-ESD), introdotta da Twitter. Identificati gli hashtag unici tra i 1000 più popolari, vengono identificati, tramite un algoritmo di Machine Learning non supervisionato applicato alle serie storiche degli hashtag (limitate attorno al picco massimo), i pattern temporali (o cluster) degli eventi. Infine, utilizzando le feature di Twitter, per ogni cluster si studia sia il processo all'origine dell'evento che l'evoluzione di quest'ultimo sulla rete.*

Abstract*: We perform a temporal analysis of the Twitter stream to investigate the evolution of unique events based on the burst of popularity of associated hashtags. We derive a classification of events according to the different patterns corresponding to the peak of the volume of exchanged message and to how these events propagate on any social network with the same characteristics as Twitter. We first provide a precise definition of unique events and correlate them to hashtags. With reference to a specific interval of time, the most popular - with respect to number of tweets- hashtags are then detected using the Seasonal Hybrid ESD (S-H-ESD) technique introduced by Twitter. After identifying the unique hashtags among the 1000 most popular, we have identified, through an unsupervised Machine Learning algorithm applied to the historical temporal series of hashtags limited around the maximum peak, the temporal patterns (clusters) of the events. Finally, using the Twitter features, for each cluster, we have studied both the process at the origin of the event and how they evolve over the network.*

1. Introduzione

Vista la semplice struttura dei dati, le funzionalità basilari offerte e la disponibilità limitata dei dataset, la piattaforma di microblogging Twitter è un punto di riferimento in ambito di ricerca per studiare particolari fenomeni [1] [2] [3] [4]. Lo studio degli eventi è un'area essenzialmente orientata all'individuazione, classificazione e analisi degli eventi significativi che si palesano sui social media, e si distingue da altre aree di ricerca per la sua complessità e per la difficoltà intrinseca di interpretare i risultati.

In questo contesto, è stata eseguita un'analisi adattando il modello proposto da Yang e Leskovec in [5], dove l'individuazione degli eventi unici avviene sulla base della popolarità (numero di tweet) degli hashtag, in uno specifico intervallo temporale. Estendendo la definizione in [6], per evento intendiamo un *fatto che causa, in un determinato periodo di tempo, un incremento sostanziale della frequenza dei messaggi (azioni degli utenti) su Twitter, tutti aventi lo stesso hashtag*; mentre per "unici" denotiamo – osservandone il trend temporale - tutti gli eventi che presentano un picco non ambiguo, e che non risultano pertanto né continui né periodici.

Per individuare gli hashtag più popolari, è stato utilizzato l'algoritmo Seasonal Hybrid ESD (S-H-ESD), una tecnica di Anomaly Detection proposta da Twitter [7] che consente di identificare i picchi sulle serie storiche come "anomalie" nell'evoluzione temporale. Basato sul test ESD generalizzato [8], S-H-ESD risulta più robusto dal punto di vista statistico – poiché utilizza metriche come la mediana e la MAD - e può essere utilizzato sia per l'identificazione delle anomalie globali che per quelle locali.

Il dataset utilizzato, contenente 19 milioni di tweet e più di 300 mila hashtag, è stato analizzato grazie ad un cluster di 8 server con Hadoop HDFS e Apache Spark. Tramite quest'ultimo framework è stato possibile estrarre i soli dati e metadati più significativi dal dataset a disposizione, generare le serie storiche degli hashtag - su scala giornaliera - e, infine, eseguire tutte le operazioni preliminari per la pulizia dei dati, anche sui testi dei singoli tweet. Dopo aver identificato gli hashtag unici tra i 1000 più popolari utilizzando la tecnica S-H-ESD, tramite un algoritmo di Machine Learning non supervisionato applicato alle serie storiche degli hashtag – chiaramente limitate attorno al picco massimo – sono stati individuati i pattern temporali (cluster) degli eventi. L'uso di un approccio non supervisionato è stato preferito a quello supervisionato per evitare ipotesi a priori sul profilo dei picchi, ma anche per evitare l'annotazione manuale del dataset al fine di distinguere i differenti tipi di evento.

Con un algoritmo di clustering sono stati individuati 5 cluster, a cui sono stati successivamente associate specifiche tipologie di evento estraendo i topic dai tweet di ogni classe grazie un Topic Model distribuito sul cluster.

E' senza dubbio possibile affermare che l'individuazione di precisi pattern sul web, come notato da [5], non è banale a causa del comportamento imprevedibile delle persone, che peraltro dipende da diversi fattori (interazione tra i singoli, in piccoli gruppi ma anche in community vaste).

Come già accennato, è stato eseguito il clustering delle serie storiche degli hashtag con un algoritmo altamente scalabile e robusto al fine di verificare l'esistenza di specifiche caratteristiche in ogni classe, anche in base alla tipologia di evento associato (determinato dai topic estratti). Infine, utilizzando le feature di Twitter [9] [10] [11] [12] [13] sui singoli cluster, sono stati studiati I processi all'origine dei profili di popolarità sul social network. Il risultato dell'analisi mostra chiaramente la dualità tra gli eventi endogeni e gli eventi esogeni.

2. Dataset

Per l'analisi è stato utilizzato un sample di Twitter in lingua italiana ottenuto dallo stream della piattaforma, sia filtrando i dati tramite una lista di stopword italiane più utilizzate, sia selezionando la lingua tramite la funzione di selezione interna. Il periodo di riferimento della collezione è compreso tra il 30 ottobre 2015 e il 3 gennaio 2016 (66 giorni totali). La Tabella 1 mostra alcune informazioni preliminari sul dataset usato.

Tabella 1. Informazioni generali dataset

Informazioni generali sul dataset	
Dimensione	21.33 GB
Numero di tweet	19000000
Numero di hashtag	377215
Tweet con hashtag	15754093

3. Tecnologie

Per il caricamento, l'estrazione e la manipolazione del sample di Twitter memorizzato sul file system distribuito di Hadoop (HDFS), è stato utilizzato il framework distribuito Spark (sia in Python che in Scala) su un cluster di 8 server, Spark RDD e Spark Dataframe come strutture dati principali.

4. Analisi degli hashtag

Per l'analisi degli eventi, ignoriamo tutti i tweet senza hashtag. Dalla Tabella 1 segue che, in questo modo, vengono rimossi meno del 20% dei tweet dal dataset. La Figura 1 mostra la percentuale di tweet con uno specifico numero di hashtag ed è interessante notare che più del 50% di essi ne contiene solo uno.

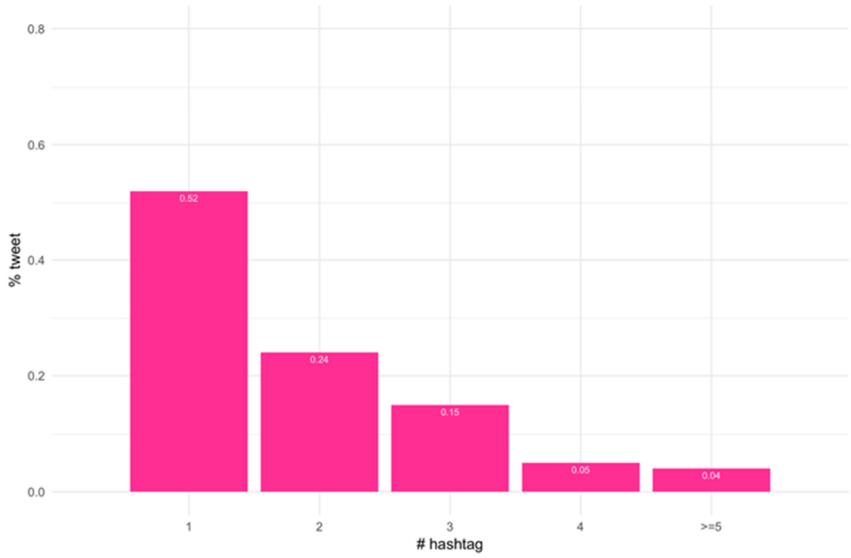

Figura 1. Distribuzione tweet per numero di hashtag

Per ogni hashtag h consideriamo il numero di occorrenze di h, e per ogni x tra 1 e il numero totale di tweet la frazione di hashtag che occorrono in almeno x tweet. La Figura 2 mostra questa relazione, che risulta statisticamente simile ad una distribuzione lognormale con parametri $(-7.99, 4.26)$ e p-value 0.71. Alla luce di quanto detto, vengono considerati i soli 1000 hashtag più frequenti.

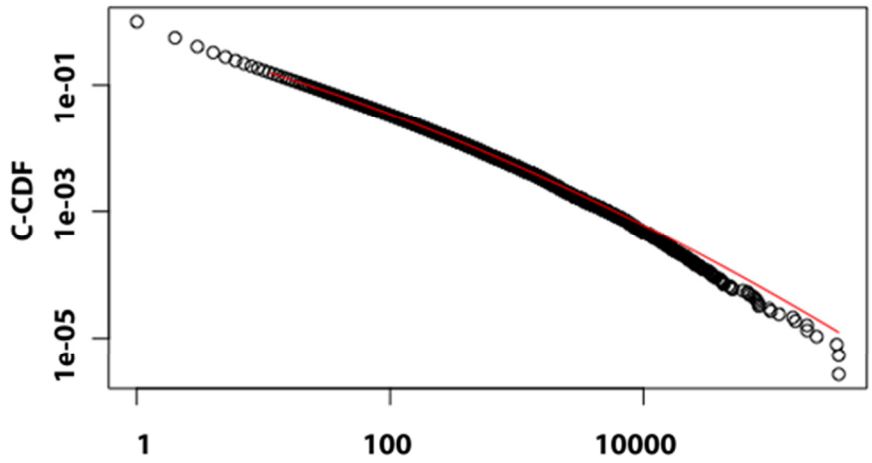

Figura 2. Numero hashtag che occorrono in almeno uno specifico numero di tweet

Siano H i 1000 hashtag più popolari. Definiamo le serie storiche di ogni h in H nell'intervallo di tempo $\tau = [0, \dots, T]$: per ogni h in H e t in τ definiamo x_{h_t} come il numero di tweet contenenti l'hashtag h pubblicati o retweettati al tempo t. In sintesi, il volume di hashtag h al tempo t. La sequenza $x_{h_0}, \dots, x_{h_T}$ è quindi la serie storica dell'hashtag h.

Poiché siamo interessati all'attività giornaliera di ogni hashtag, ogni elemento in τ rappresenta un intervallo di tempo di 24 ore, ovvero $T = 65$ (vedi sezione Dataset).

Possiamo distinguere almeno tre tipologie di serie storiche, in base ai profili emersi: le serie con profilo continuo mostrano un livello costante di attività giornaliera (vedi Figura 3); un profilo periodico è tipico delle serie storiche degli hashtag associati ad eventi che si ripetono in un intervallo fissato come, ad esempio, gli show televisivi di prima serata (vedi Figura 4); infine le serie storiche con un picco isolato rappresentano a tutti gli effetti gli hashtag associati agli eventi unici (vedi Figura 5).

In questo articolo ci si è focalizzati sull'ultima classe di hashtag poiché rappresentano quegli eventi "singolari" più interessanti da studiare. Per individuare i picchi nelle serie storiche utilizziamo l'algoritmo di Anomaly Detection Seasonal Hybrid ESD (S-H-ESD) [14] [7] basato sul test ESD generalizzato, che consente di individuare sia le anomalie locali che quelle globali [15]. Dal momento che le serie storiche degli hashtag possono esibire picchi di tipo differenti, e poiché siamo interessati ai soli picchi isolati che corrispondono alle anomalie globali, per ogni serie storica degli hashtag si ignorano tutti i picchi separati dagli altri da meno di una settimana, così come tutte le serie che non esibiscono picchi nel

proprio profilo. Vengono pertanto presi in considerazione 206 hashtag presenti in 1913470 tweet.

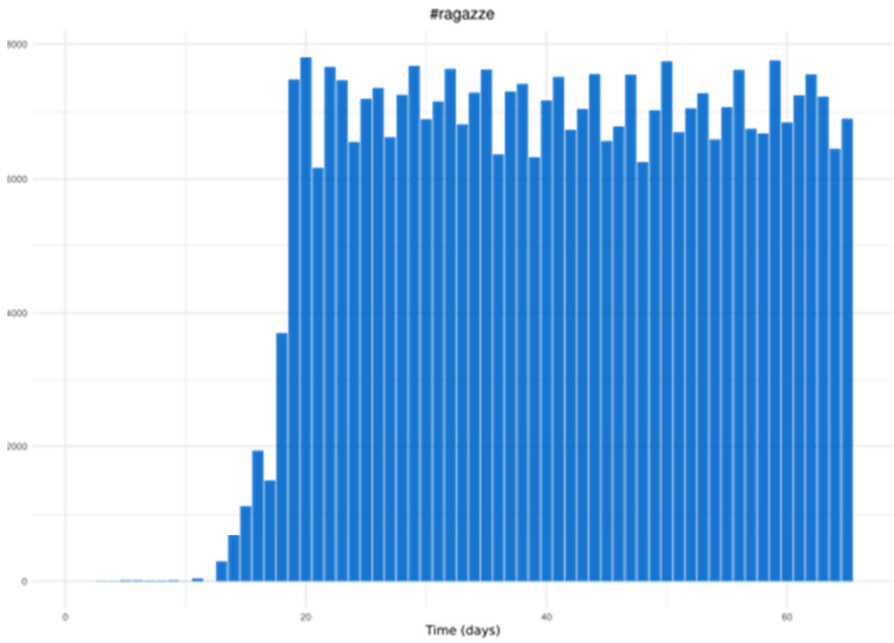

Figura 3. Serie storica hashtag "ragazze"

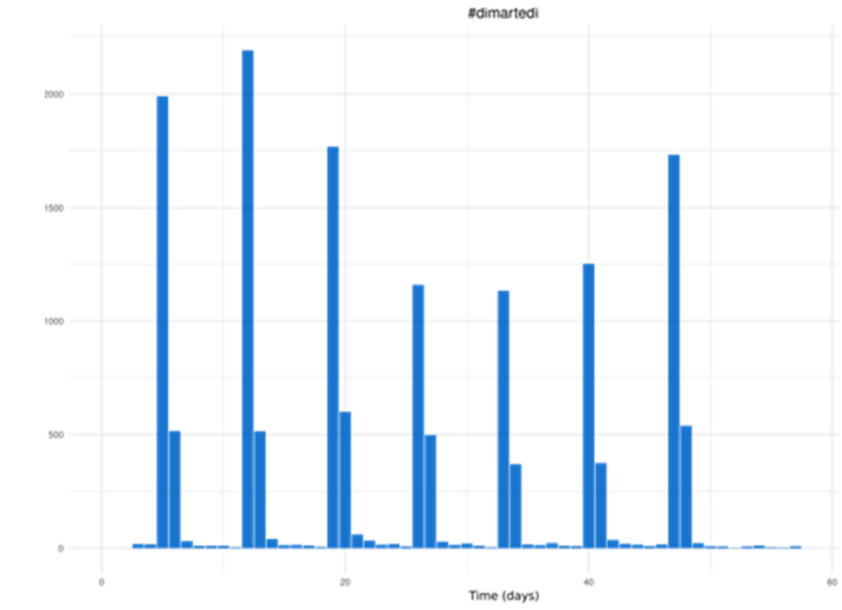

Figura 4. Serie storica hashtag "dimartedì"

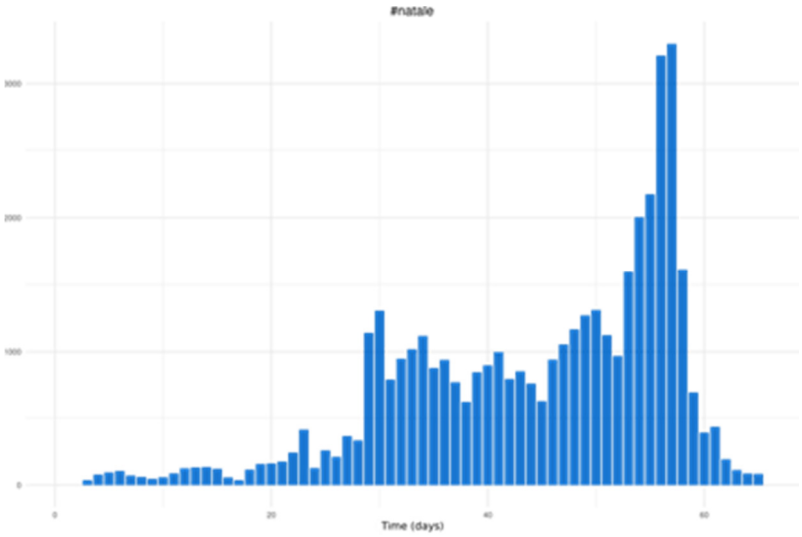

Figura 5. Serie storica hashtag "natale"

5. Periodi attivi

Un hashtag si definisce inattivo se utilizzato in meno di 20 tweet in una finestra temporale di 24 ore. Il limite definito è dovuto al rumore di fondo associato ai tweet più popolari.
La Figura 6 mostra la funzione di distribuzione cumulativa del numero di periodi attivi, mentre la Figura 7 la cumulativa della lunghezza dei periodi attivi. Dalla prima si evince che più del 90% degli hashtag risulta attivo per un massimo di 7 giorni, mentre dalla seconda che la lunghezza dei periodi attivi è sufficientemente corta: circa il 90% degli hashtag hanno periodi attivi che non superano i 10 giorni. Questo suggerisce che, come facilmente ipotizzabile, gli hashtag associati agli eventi unici risultano sporadici, occasionali e volatili

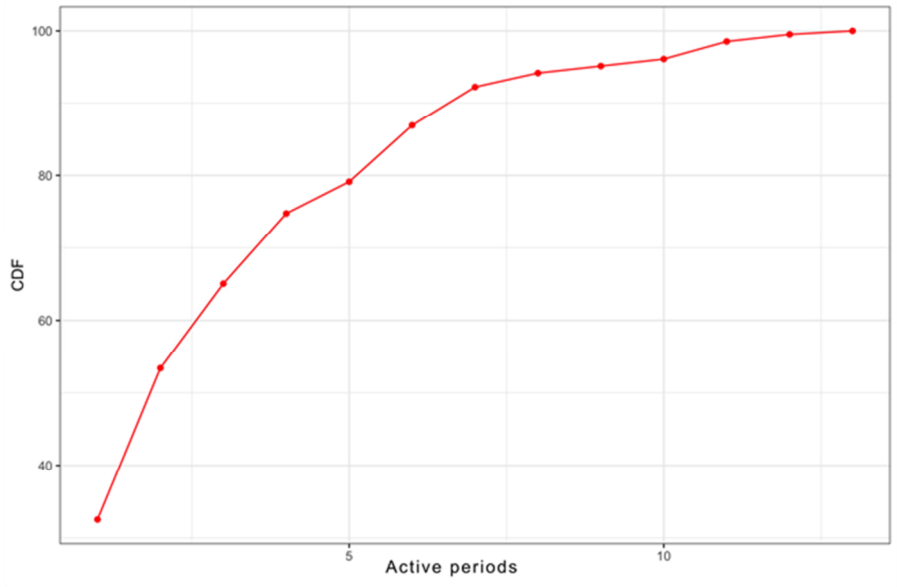

Figura 6. Cumulativa del numero di periodi attivi

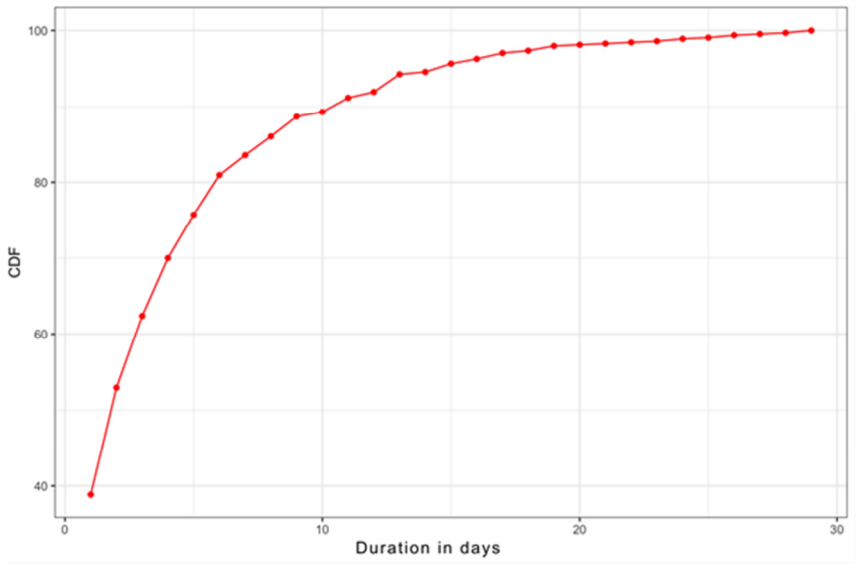

Figura 7. Cumulativa della lunghezza dei periodi attivi

6. Clustering

Classifichiamo le serie storiche degli hashtag contenenti picchi isolati con l'algoritmo di clustering K-SC, proposto in [5]. Peculiarità di questa tecnica è la sua invarianza rispetto alle operazioni di ridimensionamento e traslazione. Ne consegue che i profili con la stessa forma ma differente dimensione o posizione possono essere classificati nello stesso modo, rispetto a quanto avviene nell'algoritmo K-means [16].

Come già accennato, tutte le serie storiche oggetto della nostra analisi hanno una lunghezza pari a 65 giorni: al fine di limitare gli effetti del rumore di fondo abbiamo deciso di troncare le serie e focalizzarci sul solo intervallo di tempo attorno al picco.

Definiamo come "core" di una serie storica l'intervallo di tempo attorno al picco, e tronchiamo la serie storica eliminando i valori al di fuori del core. Sia x_p il valore massimo della serie storica, ovvero il volume al picco; T_p l'istante temporale del picco e α il valore compreso tra 0 e 1 tale che αv_p è il volume minimo che definisce il core. Il core della serie storica è l'intervallo compreso tra $T_1(\alpha)$ e $T_2(\alpha)$ dove $T_1(\alpha) < T_p$, $T_2(\alpha) > T_p$ e per tutti i t tra $T_1(\alpha)$ e $T_2(\alpha)$, il volume a t è almeno αx_p (Figura 8)

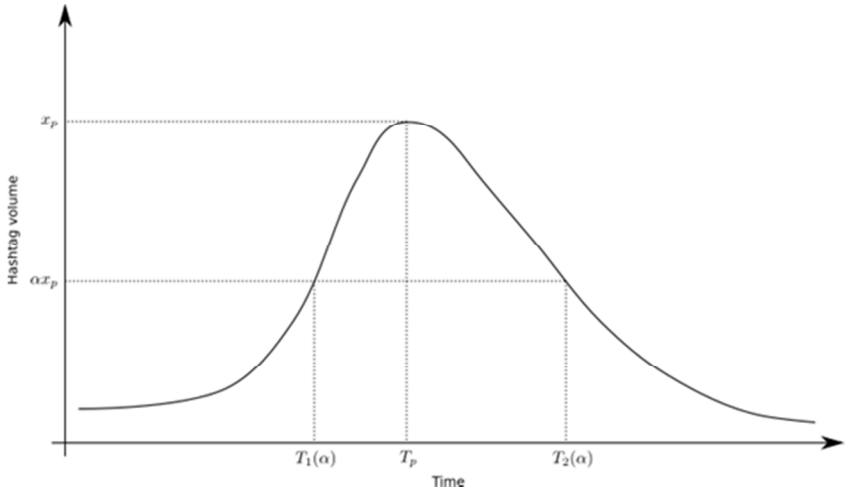

Figura 8. Tipico trend della serie storica di un hashtag

A questo punto possiamo scegliere il valore appropriato per α. Dato α, l'ampiezza del core di una serie storica è $T_2(\alpha)$ - $T_1(\alpha)$, l'ampiezza del core a sinistra è T_p - $T_1(\alpha)$, e l'ampiezza del core a destra è $T_2(\alpha)$ - T_p. La Figura 9 mostra i valori medi delle tre misure. Con valori di α più piccoli di 0.5 l'ampiezza del core a destra è maggiore rispetto a quella di sinistra, e ciò implica che la pendenza a sinistra del core è maggiore rispetto alla pendenza a destra. Abbiamo impostato la dimensione del core a 21 giorni: questo numero si ottiene osservando che in media, in due settimane, il volume minimo del picco è almeno il 12% del valore massimo (vedi Figura 8); vien aggiunta un'ulteriore settimana per gestire i valori anomali.

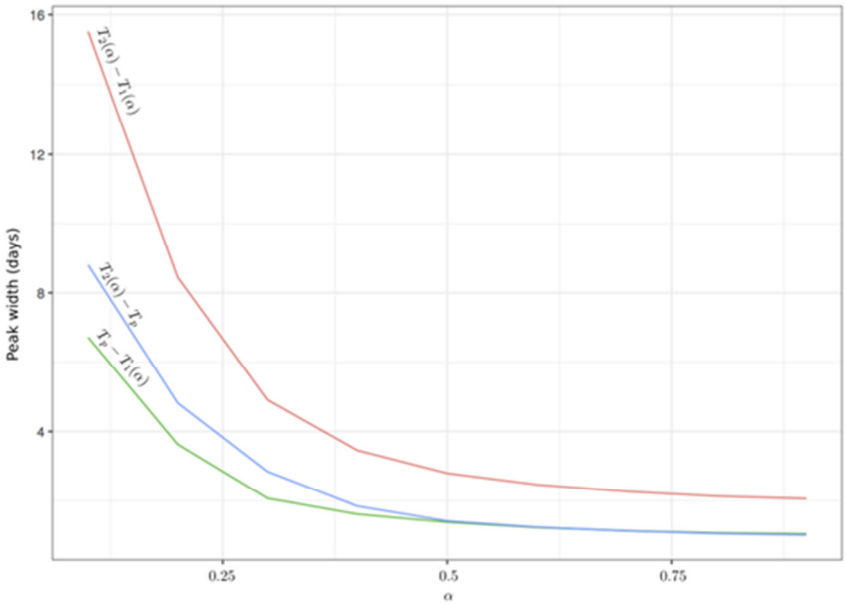

Figura 9. Ampiezza del core in funzione di α

Come noto in K-Means il numero di cluster deve essere specificato come parametro di input; e questo è valido anche per tutte le sue varianti, incluso chiaramente l'algoritmo K-SC qui utilizzato. Al fine di

scegliere il numero più appropriato di cluster, abbiamo eseguito K-SC per diversi valori di k misurando contestualmente la qualità del clustering con la metrica Average Silhouette [16].

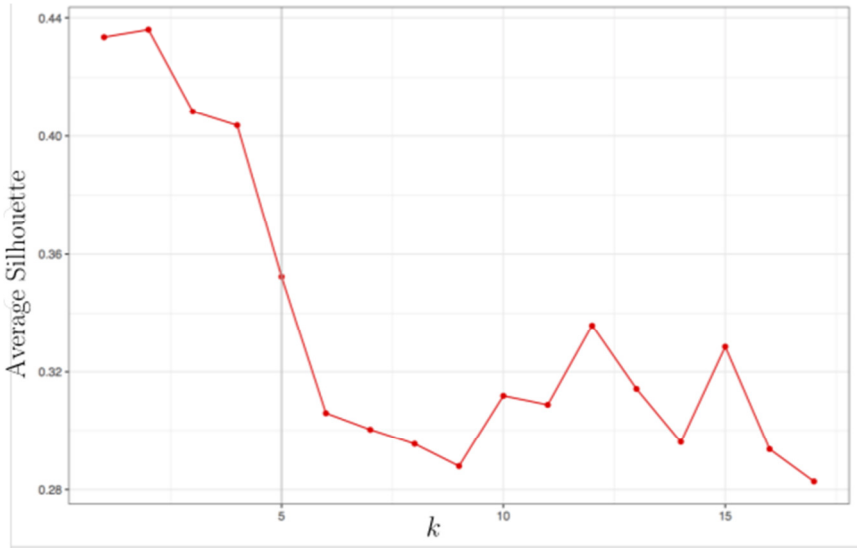

Figura 10. Average Silhouette

La Figura 10 mostra l'Average Silhouette in funzione del numero di cluster: maggiore è il valore della misura, migliore è la qualità del clustering. Nel nostro caso questo risultato si ottiene con valori piccoli di k. Abbiamo scelto quindi k = 5 poiché risulta un ottimo compromesso tra qualità e numero di cluster. Il numero di iterazioni eseguite è stato di 100.

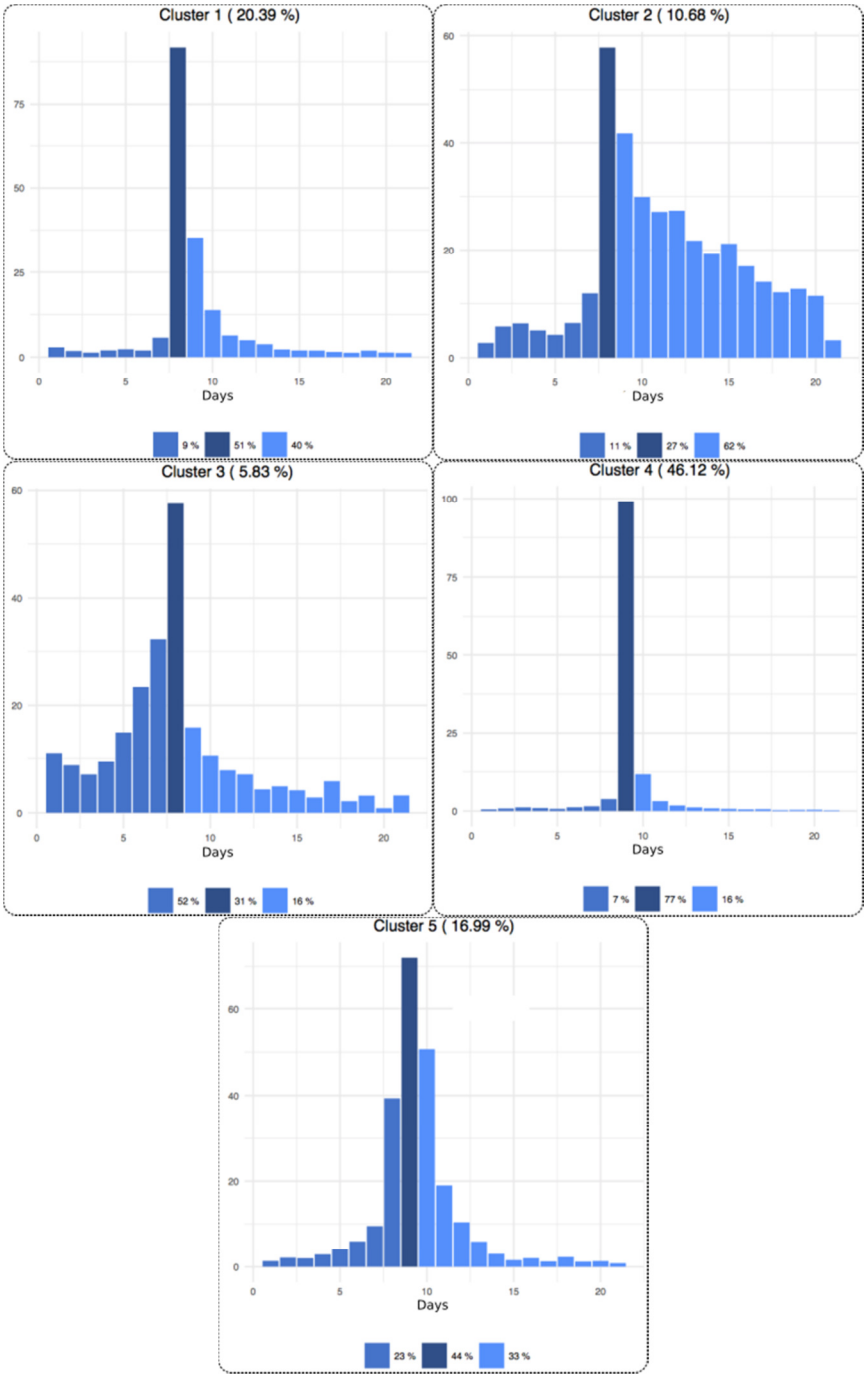

Figura 11. Risultato del clustering

La Figura 11 mostra il risultato del clustering. Per ogni cluster è indicato il numero di hashtag in esso contenuti (in percentuale), il volume al picco, a sinistra e a destra di esso (in percentuale). Il Cluster 4 contiene il maggior numero di hashtag (46.12%): il volume di tweet è concentrato in gran parte in un singolo giorno. Il pattern temporale del Cluster 1 è simile a quello del Cluster 4, se non per una coda più pronunciata: questo denota un interesse relativo della community agli argomenti trattati. Il Cluster 3 e il Cluster 4 risultano a tutti gli effetti opposti: il primo è caratterizzato da un volume relativamente elevato

dopo il picco, mentre il secondo da un volume elevato prima del picco. Infine il Cluster 5 è sostanzialmente simmetrico.

E' interessante notare che circa il 70% degli hashtag è presente in soli due cluster (1 e 4), i cui volumi sono concentrati attorno al picco.

7. Individuazione dei topic

Nel prossimo step esaminiamo i topic che caratterizzano i cinque cluster individuati.

Definiamo il singolo documento d_h per ogni hashtag h ottenuto aggregando tutti i tweet che contengono l'hashtag h. La collezione di documenti relativa ad un cluster i (con $i = 1, \dots, 5$) contiene tutti i documenti d_h tali che l'hashtag h è nel cluster i; denotiamo questa collezione con D_i. Le cinque collezioni $D_1, \dots, D_5$ vengono utilizzate come input per l'algoritmo Latent Dirichlet Allocation (LDA) al fine di estrarre i topic dei 5 cluster [17]. Il numero di topic k è un parametro dell'algoritmo, e il suo valore viene scelto eseguendo 200 iterazioni di LDA per diversi numeri di topic, al fine di individuare il numero che massimizza la log-likelihood. Valorizziamo inoltre I parametri α e β rispettivamente a $\frac{50}{k} + 1$ e 1.1.

Per ogni cluster, abbiamo quindi una lista di topic identificati dalle 5 parole più probabili. Nella Tabella 2 viene illustrata per brevità la lista relativa al solo cluster 1.

1. discorso, mattarella, presidente, italia, politica
2. grillo, italia, dittatore, blog, renzi
3. dignità, unità, papa, soldi, storia
4. renzi, italiani, italia, conferenza stampa, auguri
5. champions league, juventus, roma, italiane, mercato
6. paralizzati, mercoledi, concorso, domani, assunti
7. elezioni, prossime, partito, vince, elettorale
8. pompei, soldi, scavi, renzi, posizione
9. turchia, isis, erdogan, città, turco
10. parigi, terrorismo, morti, sicurezza, guerra
11. buonanotte, lettori, notte, libro, serata
12. renzi, ponte, acqua, messina, stretto
13. renzi, mannoia, concerto, fiorella, esclusa
14. gara, giovani, sanremo, nomi, festival
15. natale, egidio, santo, pranzo, misericordia
16. sinistra, renzi, fassina, nasce, politica
17. cinema, star, wars, forza, risveglio
18. banchetti, coraggio, scavi, piazza, renzi
19. apple, iphone, fisco, accordo, samsung
20. sanremo, giovani, solidarietà, scialpi, discriminazione
21. livorno, bilancio, società, buco, rifiuti
22. laurea, poletti, giovani, orario, italia
23. dignità, unità, papa, soldi, storia
24. bollo, targa, proposta, legge, biciclette
25. expo, milano, successo, italia, finito

Tabella 2 Topic cluster 1

Come si può evincere, i topic trattano essenzialmente eventi inattesi, spesso di carattere politico. Sono presenti inoltre topic che trattano di sport (Champions League) e intrattenimento (Festival di Sanremo). Il cluster 2 è caratterizzato da eventi inattesi come il crack di Banca Etruria. Il cluster 3 risulta invece completamente differente poiché relativo a eventi programmati come i festeggiamenti per il Natale e il Nuovo Anno, ma è presente anche Telethon, il processo Vatileaks e un evento dedicato ai libri. Il cluster 4 è caratterizzato da eventi sostanzialmente giornalieri: Black Friday, l'apertura della Porta Santa al Vaticano ed alcuni eventi sportivi. Infine, nel cluster 5 sono presenti eventi misti dal ridotto impatto sull'opinione pubblica.

8. Conclusioni

L'analisi degli eventi sui social media, la loro classificazione in base ai pattern temporali e lo studio della loro propagazione sulla rete è un'area di ricerca ampiamente studiata, spesso con tecniche e approcci totalmente differenti. A differenza di altri studi, in questo articolo è stato adottato un algoritmo di clustering degli eventi - e delle loro serie storiche - altamente scalabile. Con il dataset a disposizione, siamo riusciti ad identificare 5 differenti cluster di eventi, precedentemente selezionati tramite una tecnica di Anomaly Detection: Seasonal Hybrid ESD (S-H-ESD), proposta da Twitter e testata anche da Netflix, preferita al classico ESD poiché statisticamente più robusta.

Successivamente è stato utilizzato LDA, aggregando i tweet per ogni hashtag, con cui è stato possibile identificare le tipologie di evento associate a specifici pattern temporali, tramite un'annotazione semantica – seppur manuale - degli stessi. In particolare, il cluster 3 - poiché caratterizzato da un elevato volume di messaggi prima del picco - può essere associato agli eventi programmati. E' bene notare che questa tipologia di pattern è tipicamente associata agli eventi scatenati da fattori endogeni. Il cluster 4, invece, è associato a pattern temporali che si riferiscono agli eventi giornalieri one-shot, che risultano popolari tra gli utenti nel solo giorno in cui avvengono. Considerando l'alta proporzione di retweet, è possibile ipotizzare che siano scatenati da fattori esogeni. I cluster 1 e 2, d'altro canto, sono associati ad eventi inattesi che impattano in modo differente sulla community. Mentre nel primo caso possiamo interpretare il fenomeno come propagazione endogena, nel secondo - osservando l'elevata proporzione di tweet con url – possiamo affermare che gli eventi sono stati guidati da fattori esterni, iniettati nella rete tramite i mass media. Infine, il cluster 5 mostra un profilo simmetrico, e corrisponde a eventi misti, in cui sia i processi endogeni che quelli esogeni contribuiscono alla propagazione dell'informazione. E' chiaro che l'analisi eseguita può essere migliorata: utilizzando un dataset di dimensioni maggiori l'algoritmo K-SC potrebbe potenzialmente identificare nuovi cluster e, quindi, nuovi pattern temporali. Inoltre potrebbe essere interessante studiare la propagazione degli eventi analizzando il Follow Graph di ogni utente [1], così come

utilizzare algoritmi di Natural Language Processing e Named Entity Recognition assieme a LDA per migliore il risultato del topic model.

Bibliografia

[1] G. Amati, S. Angelini, G. Gambosi, G. Rossi, P. Vocca e G. Marcone, Moving Beyond the Twitter Follow Graph, Lisbon: KDIR 2015 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015), Volume 1, Lis- b, 2015.

[2] G. Amati, S. Angelini, F. Capri, G. Gambosi, G. Rossi e P. Vocca, Twitter Temporal Evolution Analysis: Comparing Event and Topic Driven Retweet Graphs, Funchal: {BIGDACI} 2016 - Proceedings of the International Conference on Big Data Analytics, Data Mining and Computational Intelligence, Volume 1, 2016.

[3] G. Amati, S. Angelini, F. Capri, G. Gambosi, G. Rossi e P. Vocca, Modelling the temporal evolution of the retweet graph., IADIS International Journal on Computer Science and Information Systems, 2016.

[4] G. Amati, S. Angelini, F. Capri, G. Gambosi, G. Rossi e P. Vocca, On the Retweet Decay of the Evolutionary Retweet Graph, Venice: Smart Objects and Technologies for Social Good: Second International Conference, GOODTECHS 2016, 2017.

[5] J. Yang e J. Leskovec, Patterns of Temporal Variation in Online Media, New York: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)., 2011.

[6] W. Dou, X. Wang, D. Skau, W. Ribarsky e M. X. Zhou, LeadLine: Interactive visual analysis of text data through event identification and exploration, 2012 IEEE Conference on Visual Analytics Science and 593 Technology (VAST), 2012.

[7] S. Kelly e K.Ahmad, Propagating Disaster Warnings on Social and Digital Media, Intelligent Data Engineering and Automated Learning -- IDEAL 2015, 2015.

[8] B. Rosner, Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics, 1983.

[9] M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi e M. Zaharia, Spark SQL: Relational Data Processing in Spark, Melbourne: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 2015.

[10] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan e X. Li, Comparing Twitter and Traditional Media Using Topic Models, Berlin: Advances in Information Retrieval, 2011.

[11] S. Asur, B. A. Huberman, G. Szabo e C. Wang, Trends in Social Media : Persistence and Decay, CoRR, 2011.

[12] J. Lehmann, B. Goncalves, J. J. Ramasco e C. Cattuto, Dynamical Classes of Collective Attention in Twitter, Lyon: Proceedings of the 21st International Conference on World Wide Web, 2012.

[13] W. Guan, H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Cao e X. Yang, Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events, Physica A: Statistical Mechanics and its Applications, 2013.

[14] A. Kejariwal, Introducing practical and robust anomaly detection in a time series, https://blog.twitter.com/engineering/en_us/a/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series.html, 2015.

[15] B. Rosner, Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics, 1983.

[16] L. Kaufman e P. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Number Book 59 in Wiley Series in Probability and Statistics. Wiley-Interscience., 2005.

[17] D. Blei, A. Ng e M. Jordan, Latent Dirichlet Allocation, J. Mach. Learn. Res., 2003.

[18] S. Kelly e K. Ahmad, Propagating Disaster Warnings on Social and Digital Media, Cham: Intelligent Data Engineering and Automated Learning – IDEAL 2015,, 2015.