

Tv e Social WEB

La seconda vita del contenuto televisivo

Roberto **Del Pero**, Claudio **Schifanella**, Luca **Vignaroli**
Rai - Centro Ricerche e Innovazione Tecnologica

Ruggero **Pensa**, Maria Luisa **Sapino**
Università degli Studi di Torino, Dipartimento di Informatica

1. INTRODUZIONE

Negli ultimi anni il modo in cui gli utenti fruiscono del contenuto televisivo è radicalmente cambiato. Il processo di digitalizzazione dei contenuti e delle trasmissioni ha permesso una notevole evoluzione dell'ecosistema televisivo verso nuove forme di fruizione, navigazione e interazione. Possiamo inoltre notare profondi cambiamenti anche nelle caratteristiche tipiche dell'ambiente all'interno del quale i contenuti sono fruiti: sempre più spesso gli utenti guardano i contenuti televisivi utilizzando al tempo stesso dispositivi mobili come smartphone e tablet per interagire, commentare e approfondire i programmi che vengono trasmessi dai broadcaster [5][3]. La popolarità dei social network ha modificato l'ecosistema Web verso un ambiente più collaborativo e dove l'importanza dei contenuti generati dall'utente diventa sempre maggiore. L'analisi di questi contenuti certamente non può rappresentare in senso stretto un campione statisticamente rappresentativo della società, ma certamente deve essere tenuto in considerazione da tutti gli attori per poter cogliere tendenze, popolarità e opinioni di una parte di utenti che non può essere trascurata.

Tipicamente il contenuto televisivo prodotto e messo in onda dal broadcaster viene arricchito con descrizioni in linguaggio naturale e metadati e, successivamente, memorizzato in maniera statica in archivio per successivi riutilizzi. Il broadcaster può, inoltre, offrire online i contenuti TV attraverso un proprio portale Web; operazioni altrettanto tipiche

L'articolo descrive i risultati ottenuti nell'ambito della collaborazione tra il Centro Ricerche RAI e il Dipartimento di Informatica dell'Università degli Studi di Torino nella definizione di un modello di integrazione di dati eterogenei provenienti da differenti basi di conoscenza (archivi televisivi, EPG, audience, social networks, ecc) che contribuiscono alla definizione della "seconda vita" del contenuto televisivo, a partire dalla sua produzione, la successiva messa in onda e proseguendo con i commenti e contenuti generati sul web dagli utenti.

Il modello proposto permette di rappresentare attraverso un unico grafo di conoscenza le differenti entità e relazioni che emergono dal dominio televisivo analizzato.

sono l'analisi dell'audience allo scopo di quantificare l'indice di gradimento della trasmissione oppure l'analisi degli accessi al proprio sito Internet. Cosa succede in seguito? Molti dei programmi messi in onda, nella fattispecie quelli più interessanti, vengono ripubblicati in rete, in tutto o in parte, dagli utenti stessi (ad esempio su YouTube), durante e dopo la messa in onda i contenuti televisivi vengono commentati dagli utenti sui principali portali di social network, come Facebook e Twitter, dando vita anche ad intensi dibattiti sulla rete.

Per i contenuti TV inizia quindi una *seconda vita* durante la quale il contenuto audio/video viene guardato, "taggato", "apprezzato", commentato e condiviso più e più volte, trasformandosi in una "calamita" in grado di attrarre utenti del Web e diventando in sostanza un **Oggetto Sociale** che veicola emozioni, pareri, argomenti di discussione: l'analisi di tali contenuti può aiutare i diversi attori a quantificare la popolarità dei personaggi e dei programmi televisivi, i trend e il gradimento del pubblico.

L'attività di collaborazione del **Centro Ricerche RAI** con l'**Università degli Studi di Torino** oggetto

di questo articolo mira alla definizione di un framework di integrazione di dati eterogenei provenienti da sorgenti sia tradizionali che sociali. Il sistema definito permette di rappresentare in un unico modello unificato la conoscenza proveniente dalle diverse sorgenti analizzate, abilitando innovativi strumenti di ricerca e analisi inter-sorgente.

2. UN FRAMEWORK PER L'INTEGRAZIONE E L'ANALISI DI DATI SOCIALI

In questa sezione viene mostrata l'architettura del framework che permette l'integrazione di sorgenti eterogenee, sia sociali che tradizionali, in un'unica knowledge base. Essa si basa sulla definizione di un grafo di conoscenza che consente di rappresentare i concetti e le relazioni tra di essi attraverso un unico modello.

In figura 1 è rappresentata l'architettura generale del framework di integrazione. Esso è composto da tre livelli principali: il **Source Processing Layer**, il **Knowledge Graph Layer** e il **Knowledge Query and Analysis Layer**.

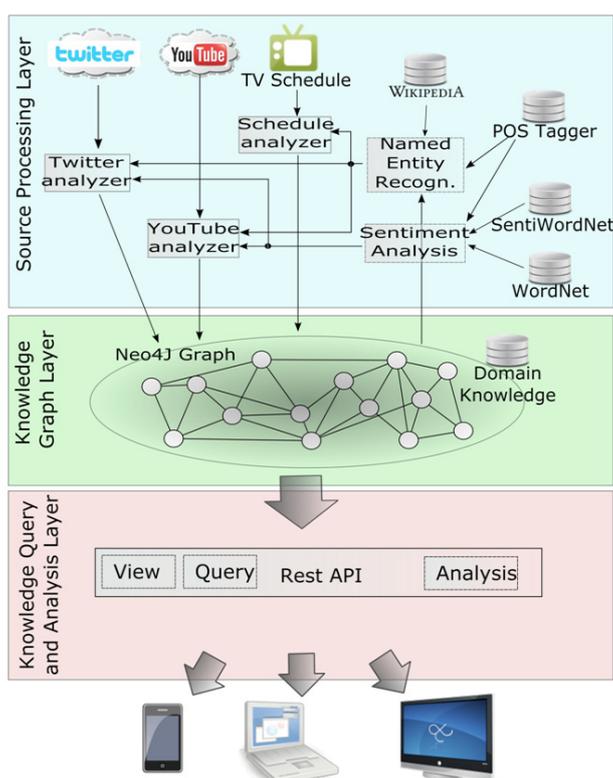


Fig. 1 – Architettura del sistema di integrazione e analisi

2.1 IL SOURCE PROCESSING LAYER

Il primo livello dell'architettura ha il compito di collezionare tutti i dati dalle sorgenti in ingresso, dati che saranno successivamente memorizzati nel grafo di conoscenza, cuore del framework proposto in questo articolo. Esso accede continuamente ad un numero predefinito di sorgenti, siano esse siti web, social networks, metadati che descrivono i programmi TV, e, attraverso una serie di moduli di analisi, estrae le unità informative che andranno a formare i nuovi nodi e relazioni all'interno del grafo di conoscenza. Il modulo denominato *Schedule Analyzer* ha il compito di leggere le informazioni del palinsesto televisivo e di orchestrare i diversi estrattori in base, ad esempio, ai programmi in onda.

Ogni sorgente è associata ad un estrattore che ha come input lo stream di dati generati dalla sorgente stessa (ad esempio i *tweet* associati al talk show televisivo *Ballarò*, o i *post* e i *commenti* presenti nella pagina *Facebook* del programma) e produce come risultato l'insieme dei concetti (personaggi, luoghi, eventi, utenti, emozioni, ecc) unitamente alle relazioni che li legano. Come vedremo nella sezione successiva, il grafo di conoscenza rappresenta un modello temporale dei dati provenienti dalle diverse sorgenti analizzate e quindi permette di memorizzare al suo interno l'evoluzione dei fenomeni osservati. Per fare ciò, ogni analizzatore divide il flusso di dati in finestre temporali e, attraverso una fase di analisi, estrae l'insieme di entità più rilevanti, collegandole tra loro attraverso opportune relazioni, in modo da formare una nuova porzione del grafo, annotata temporalmente, che nella sezione successiva verrà identificato come *Oggetto Sociale*. È importante notare come la dimensione della finestra temporale scelta può variare in funzione del contesto di analisi: è possibile decidere, ad esempio, di diminuire l'ampiezza della finestra temporale durante la messa in onda del programma per poter modellare i fenomeni sociali con una granularità maggiore, rispetto ai momenti in cui lo stesso programma non è in onda e quindi è, presumibilmente, meno commentato.

Per la fase di estrazione delle informazioni dal flusso di dati, il sistema si affida ad una serie di moduli

condivisi tra i vari estrattori che sono responsabili dell'esecuzione delle operazioni più comunemente utilizzate in questo ambito. Tra di essi possiamo trovare i moduli che svolgono i classici task dei sistemi di *Information Extraction*, quali: strumenti di analisi del testo, riconoscimento delle entità, analisi dei sentimenti.

2.1.1 Il riconoscimento delle entità

Il modulo denominato *Named-Entity Recognizer (NER)* ha il compito, dato in input un *tweet* o un commento su *YouTube*, di riconoscere le entità nominate (persone, luoghi, eventi) all'interno del testo. Possiamo individuare due fasi principali: individuazione delle entità e disambiguazione delle entità [9]. La prima fase è stata implementata attraverso l'uso sinergico di un modulo di analisi del testo, *Freeling POS Tagger* [13], coadiuvato dall'uso di *Wikipedia* come base di conoscenza strutturata. Il primo si occupa di riconoscere all'interno del testo le parti che fanno riferimento a nomi propri o comuni (es: "#barackobama", "Matteo Renzi", "Milano"), eliminando verbi, aggettivi, articoli, ecc., mentre *Wikipedia* viene utilizzata per ottenere un unico riferimento alle entità a partire da chiavi di ricerca differenti (ad esempio, "Renzi", "Matteo Renzi", "#renzi").

La seconda fase del modulo *NER* si occupa dell'operazione più importante che compete a questa unità di analisi: la *disambiguazione*. È importante notare, infatti, che ad una singola chiave di ricerca spesso corrispondono più entità candidate. Questo può avvenire sia per casi di omonimia ("Paolo Rossi" può riferirsi ad un attore o ad un calciatore) o perché il nome del personaggio non è completamente specificato ("Giannini" può riferirsi a Massimo Giannini, l'attuale conduttore del talk show televisivo *Ballarò*, o all'attore Giancarlo Giannini). La maggior parte delle soluzioni proposte in letteratura si basano sul modello *Bag of Words*, nel quale il processo di disambiguazione si basa sul calcolo della similarità tra i termini che precedono o seguono le entità da disambiguare. Come dimostrato in letteratura, queste tecniche sono inefficaci negli scenari, come quello considerato in questo articolo, caratterizzati da testi brevi e non sempre strutturati, come *tweet*

o commenti su *YouTube* o *Facebook*. Per questa ragione, il modulo di disambiguazione proposto in questo articolo fa uso dei metadati che descrivono i programmi televisivi per cercare di stabilire quale tra i candidati restituiti dalla ricerca su *Wikipedia* rappresenta l'entità che l'utente intendeva menzionare. Più in dettaglio, per ogni programma televisivo, è stato creato una mappatura con le categorie di *Wikipedia* all'interno delle quali ogni voce è catalogata: il sistema quindi, tra i possibili candidati, prediligerà quelli le cui categorie saranno più simili a quelle associate al programma televisivo che si sta esaminando. Grazie a questo algoritmo, la chiave di ricerca "Paolo Rossi", menzionata all'interno di uno stream di *Twitter* de "La domenica sportiva" restituirà come personaggio più probabile il calciatore; allo stesso modo, la ricerca del termine "Giannini" durante l'analisi del talk show "Ballarò" riconoscerà come entità più probabile l'attuale conduttore "Massimo Giannini". Inoltre, il sistema può essere proattivamente addestrato attraverso un sistema di supervisione guidato da un utente esperto del dominio.

2.1.2 La sentiment analysis

Il modulo di *analisi dei sentimenti* è responsabile dell'estrazione dei *valori di polarità* e delle *emozioni* dalle varie sorgenti analizzate. Dopo una fase preliminare di estrazione dei lemmi, effettuata sempre con l'utilizzo di *Freeling POS tagger*, il sistema utilizza il vocabolario di *SentiwordNet* [1], eventualmente coadiuvato dall'utilizzo di *MultiwordNet* [14] per la traduzione, per estrarre il valore di *polarità* (*positiva*, *negativa* o *neutra*) eventualmente associato ad ogni singolo termine. Una funzione di aggregazione permette quindi di calcolare il valore complessivo della polarità dell'insieme dei dati prodotti dagli utenti nella finestra temporale esaminata. Un algoritmo analogo viene utilizzato per l'estrazione delle *emozioni*, basato in questo caso sulla tassonomia fornita da *WordNet Affect* [18]. In questo modo ogni estrattore è in grado di associare ad ogni blocco di dati facente riferimento ad una specifica finestra temporale di uno specifico programma una rappresentazione numerica delle opinioni degli utenti attivi sui social networks.

2.2 IL KNOWLEDGE GRAPH LAYER

Il successivo livello (*Knowledge Graph Layer*) rappresenta il cuore di tutto il sistema e permette di rappresentare i concetti e le relazioni estratti dalle diverse sorgenti eterogenee in un modello di conoscenza unificato e omogeneo. Il modello proposto è una sintesi di consolidate teorie sociali proposte negli anni passati in differenti campi, quali le *scienze cognitive* [2][10][8][17], la *filosofia del linguaggio* [15] e le *ontologie sociali* [6][16].

Sulla base delle teorie appena menzionate, possiamo quindi individuare tre entità principali, corrispondenti alle tre tipologie di nodi che sono alla base del grafo proposto: gli utenti che agiscono, i *soggetti*, il risultato delle loro azioni, gli *oggetti sociali* e la rappresentazione degli oggetti da loro menzionati, i *concetti*. Ogni atto (o insieme di atti) che un utente effettua nei social network analizzati e che produce un'evidenza sociale è rappresentato all'interno del grafo da un *oggetto sociale*. Un *oggetto sociale* permette di evidenziare una relazione, all'interno di uno specifico intervallo temporale, tra un insieme di concetti, siano essi personaggi, programmi TV, luoghi, eventi, emozioni, ecc. In questo modo è possibile rappresentare, ad esempio, che il giornalista Ferruccio De Bortoli è tra i personaggi più nominati o che la paura è una delle emozioni prevalenti tra quelle veicolate dagli utenti di *Twitter*, durante un particolare intervallo temporale, rispetto ad una determinata puntata del talk show "Ballarò".

La struttura del modello prevede quindi la definizione della tipologia di relazioni che permettono di legare *soggetti*, *oggetti sociali* e *concetti*. In particolare, un insieme di utenti (*soggetti*), che commentano un programma televisivo, *supportano* la creazione di un oggetto sociale; un oggetto sociale è la *rappresentazione* di una istanza sociale di uno o più concetti in un particolare contesto (ad esempio, un video può essere una rappresentazione di un evento sportivo). I concetti e, più in generale, le entità dello stesso tipo possono essere connesse tra loro al fine di creare *strutture* complesse: ad esempio, un commento ed un video sono due *oggetti sociali* che possono essere connessi da una relazione strutturale per

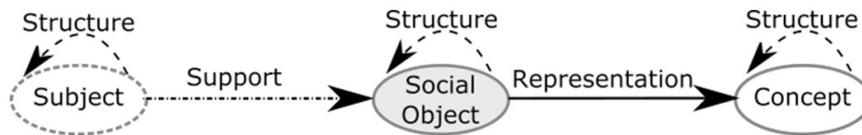


Fig. 2 – Struttura generale del modello a grafo

indicare che il primo è parte del secondo. Una tassonomia che classifica geograficamente i concetti che esprimono dei luoghi è connessa attraverso la stessa tipologia di relazione. La figura 2 riassume la struttura del grafo.

Infine, è importante notare che gli oggetti sociali sono per loro natura transienti ed evolvono nel tempo. Per questo motivo, ogni oggetto sociale può essere caratterizzato attraverso un nodo temporale che esprime un intervallo di validità, connesso attraverso una relazione di tipo temporale.

La figura 3 nella pagina seguente mostra un caso reale della struttura del grafo di conoscenza utilizzato all'interno del sistema proposto. Si riferisce al risultato dell'analisi di due stream di dati provenienti da *Twitter* e da *YouTube* relativi allo stesso programma TV (la telecronaca della finale dei 100 metri durante le Olimpiadi di Londra 2012).

Come si può notare, i vari estrattori hanno permesso di individuare i personaggi più nominati dagli utenti all'interno di *Twitter* e dei commenti ad un video *YouTube*, il concetto relativo all'evento commentato e le emozioni veicolate nei commenti. Per ogni sorgente, è stato creato un oggetto sociale complesso (*TweetSet* in un caso, *YouTube Video* nell'altro) che permette di legare gli utenti, classificati per tipologia (*active users, supporters, broadcasters*) ai concetti più rilevanti (*Usain Bolt, Asafa Powell* e altri).

Come accennato precedentemente, è possibile notare come gli oggetti sociali acquistano una validità temporale grazie alla presenza di una relazione temporale con un nodo che esprime un intervallo di tempo: in questo modo, dato un intervallo temporale di interesse, il **livello di query** introdotto nella prossima sezione sarà in grado di individuare, analizzare e restituire in maniera efficiente solamente i dati che sono validi in tale intervallo temporale.

I concetti che sono rappresentazione comune di oggetti sociali provenienti dall'analisi di sorgenti differenti permettono, inoltre, di creare un *ponte semantico* tra i diversi media, abilitando quindi meccanismi di *analisi inter-sorgente*.

Infine, la figura 3 permette di notare come il grafo memorizzi al suo interno solamente le relazioni più significative che emergono dall'analisi svolta dagli estrattori. Infatti, rappresentare ogni singola relazione e concetto, anche quelle meno frequenti, renderebbe non gestibile il modello nella successiva fase di analisi, senza apportare benefici evidenti ai risultati finali.

Per l'implementazione del grafo di conoscenza descritto è stato scelto il database NoSQL **Neo4j** [11], che rappresenta una tra le tecnologie leader del settore. **Neo4j** permette di gestire agevolmente grandi quantità di nodi e relazioni e offre una comoda interfaccia di gestione basata su meccanismi standard (*REST*). Il linguaggio dichiarativo di interrogazione **Cypher** [4] rappresenta un efficace strumento di query basato sui meccanismi standard di gestione dei grafi.

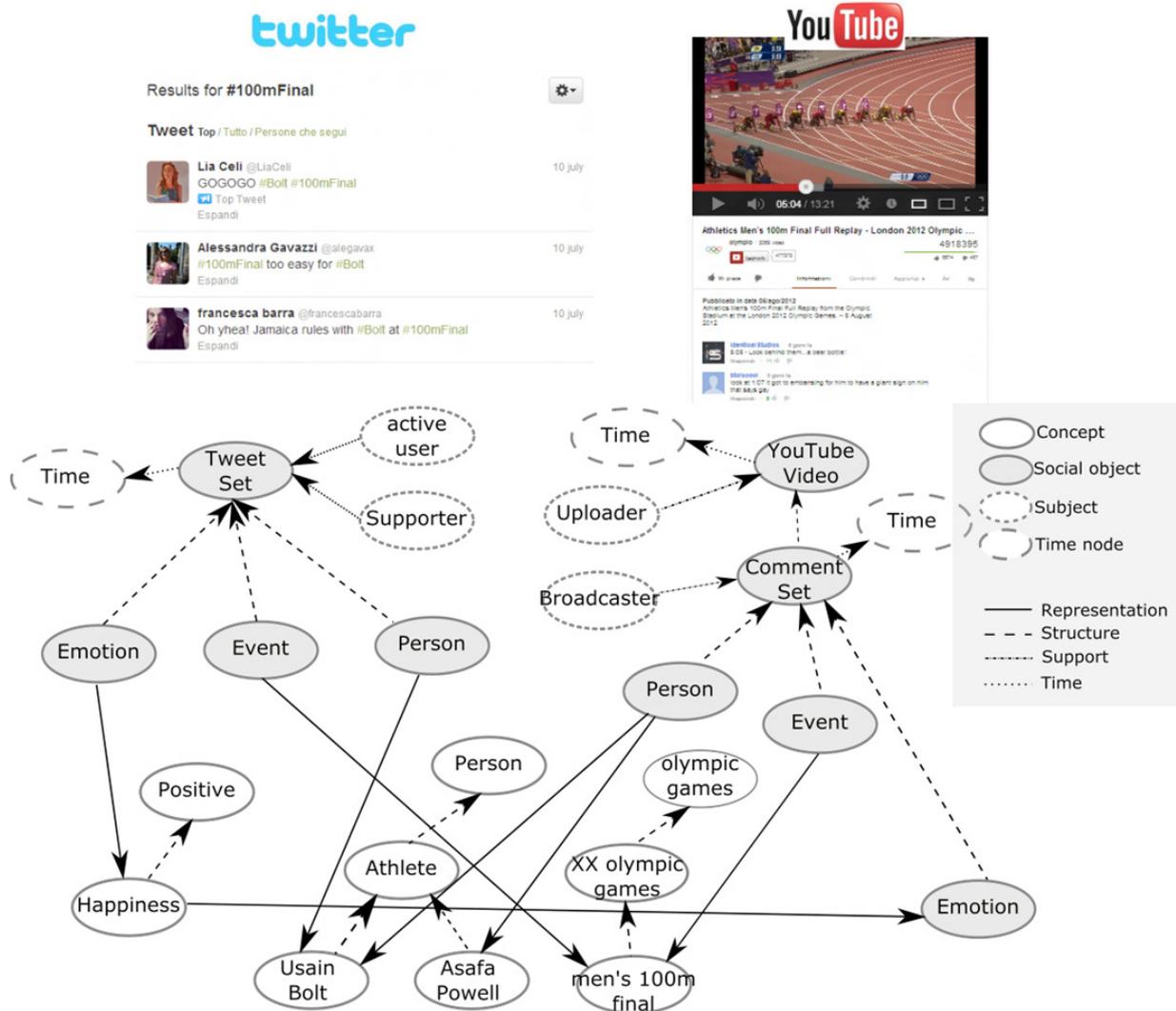


Fig. 3 – Un esempio di estrazione e integrazione dei dati provenienti da due differenti social network (Twitter e YouTube)

2.3 IL QUERY AND ANALYSIS LAYER

Il terzo livello del sistema proposto consiste in una serie di componenti responsabili dell'*interrogazione*, *navigazione* e *analisi*. Il **modulo di query** permette di estrarre sottografi a partire dal grafo di conoscenza principale, in base a criteri e vincoli stabiliti dall'utente. Ogni sottografo estratto può essere considerato come una *vista parziale* dell'intero grafo contenente solo i nodi e le relazioni rilevanti per l'utente. Il **modulo di analisi** mette a disposizione dell'utente finale una serie di strumenti basati su algoritmi di data mining in grado di estrarre pattern ricorrenti, popolarità dei personaggi, emozioni emergenti e altro. Esso può agire direttamente sull'intero grafo o utilizzare le viste estratte con il *modulo di query* o ulteriori viste che possono essere interpretate come differenti strutture dati tipiche del data mining come *matrici* o *tensori*.

3. UN CASO DI STUDIO: I TALK SHOW POLITICI

In questa sezione viene descritta una particolare sperimentazione del framework proposto, basata su dati relativi alle puntate del programma RAI "Ballarò" andate in onda tra il 2 ottobre e il 27 novembre 2012. Il periodo selezionato è di particolare interesse perché molto ricco per quanto riguarda l'attività politica: tra gli eventi rilevanti possiamo citare le

elezioni amministrative in alcune delle principali regioni italiane (Sicilia, Lazio e Lombardia), le allora imminenti elezioni politiche, la recessione, la crescita dell'interesse e del dibattito intorno al *Movimento 5 Stelle*. Sono state considerate due sorgenti di tipo sociale: *Twitter* e *YouTube*. Per ogni puntata sono stati collezionati tutti i *tweet* contenenti le chiavi di ricerca *#ballarò* (hashtag ufficiale del programma) o *@RaiBallaro* (profilo *Twitter* ufficiale del programma). I commenti estratti da *YouTube* sono stati ottenuti, attraverso un'euristica di ricerca sui video correlati, a partire dal titolo del programma e dalla data di messa in onda ricavata dal modulo di analisi del palinsesto. Qui di seguito sono riportati alcuni casi di studio che aiutano a capire le potenzialità del framework.

3.1 LA SOCIAL CENTRALITY

Il primo esempio considerato è lo studio dell'importanza (in termini di centralità dei nodi) dei personaggi (politici, conduttori, opinionisti e così via) durante la messa in onda del programma "Ballarò". Per fare ciò, abbiamo considerato tutti i personaggi che sono stati ritenuti rilevanti da ogni estrattore (*Twitter* e *YouTube* in questo caso) durante il periodo di osservazione, costruendo, attraverso il **modulo di query**, un grafo come quello riportato in figura 4.

Ogni grafo è stato costruito aggiungendo un arco

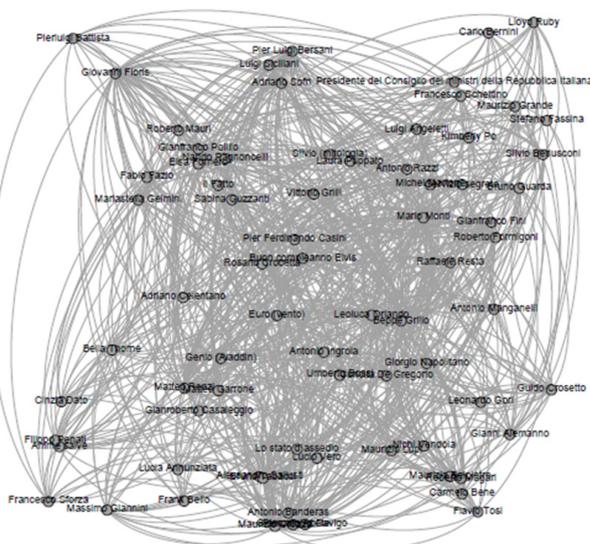


Fig. 4 – Grafo estratto dal modulo di query utilizzato per il calcolo della rilevanza sociale dei personaggi all'interno del caso di studio proposto

tra due personaggi se nell'intervallo temporale considerato esiste un percorso che li lega passando per un nodo di tipo *concetto* che rappresenti una puntata della trasmissione considerata. Dal momento che ad ogni episodio corrispondono molteplici finestre temporali di analisi, il *modulo di query* ha preventivamente aggregato tutti gli *oggetti sociali* validi nell'arco temporale considerato. È importante notare, inoltre, che i percorsi trovati possono anche coinvolgere sorgenti differenti, implementando quindi quello che, nella precedente sezione, abbiamo chiamato *analisi inter-sorgente*.

Dato il grafo sociale costruito secondo la procedura appena descritta, il calcolo della *betweenness centrality* [7][12] di ogni singolo nodo (cioè la frazione di percorsi minimi da ogni coppia di vertici

che passano anche dal nodo considerato), permette di ottenere la lista ordinata della *centralità* del concetto all'interno del grafo di conoscenza.

I risultati riportati in tabella 1 evidenziano che sia il comico Maurizio Crozza che l'ex-comico, ora leader del *Movimento 5 Stelle*, Beppe Grillo risultano essere i personaggi più discussi nel periodo di analisi, sebbene il secondo non abbia mai partecipato direttamente al programma come ospite. Seguono il Presidente del Consiglio dell'epoca, Mario Monti, ed il conduttore della trasmissione, Giovanni Floris, ma anche la *centralità* di personaggi che sarebbero diventati nuovi leader politici, come Matteo Renzi, è tra le più alte.

Tab. 1 – Valori di rilevanza sociale ottenuti con il calcolo della *betweenness centrality* (prime dieci posizioni)

Rank	Personaggio	Betweenness centrality
1	Maurizio Crozza	0.1710
2	Beppe Grillo	0.1710
3	Mario Monti	0.1305
4	Giovanni Floris	0.0954
5	Pier Luigi Bersani	0.0868
6	Matteo Renzi	0.0723
7	Gianfranco Polillo	0.0259
8	Silvio Berlusconi	0.0236
9	Bruno Tabacci	0.0234
10	Gianfranco Fini	0.0217

3.2 STUDIO DELLA POPOLARITÀ

Il secondo esperimento proposto consiste nel calcolo dell'evoluzione della popolarità dei personaggi nelle singole puntate considerate. Il concetto di popolarità rappresenta un tema molto dibattuto in letteratura ed in questo esperimento viene legato alla percentuale di menzioni del personaggio fatte nell'arco temporale di interesse dagli utenti delle due sorgenti analizzate.

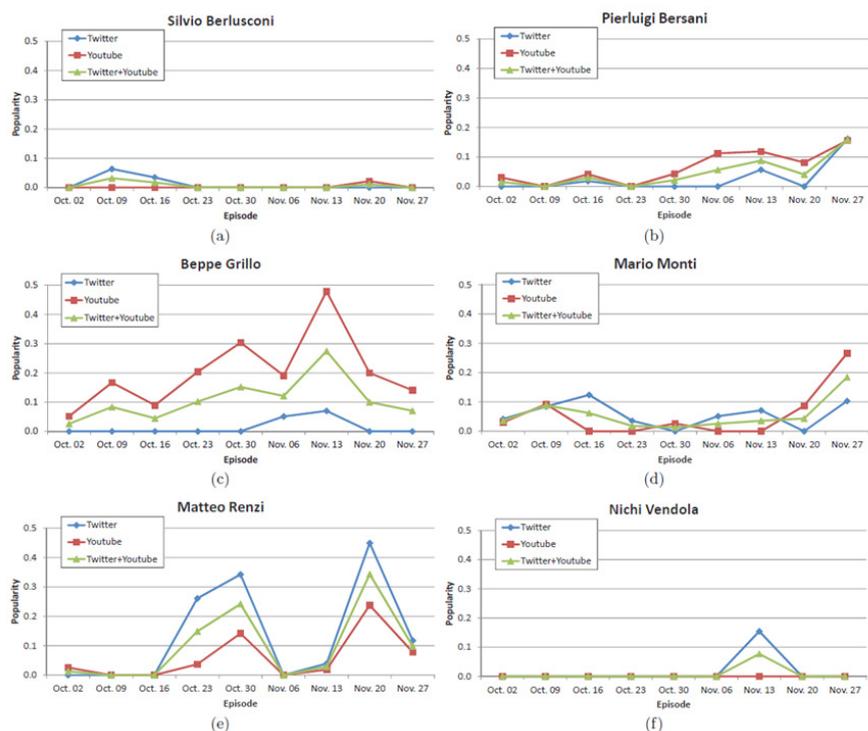
E' importante notare che questa informazione, in forma non aggregata, è calcolata da ogni estrattore. Viene memorizzata nel grafo di conoscenza all'interno dell'oggetto sociale complesso come peso della relazione di rappresentazione che lega l'oggetto sociale stesso al concetto rappresentato, frutto dell'analisi di ogni singola finestra temporale. Data una singola sorgente, risulta quindi semplice per il modulo di analisi, aggregare proporzionalmente tutti i pesi di tali archi per ottenere, per ogni sorgente la popolarità relativa di ogni personaggio. A questo punto la popolarità complessiva intersorgente di ogni personaggio è ottenuta attraverso una combinazione lineare dei singoli valori:

$$w(i)_{all} = \alpha \cdot w(i)_{tw} + (1 - \alpha) \cdot w(i)_{yt}$$

dove $w(i)_{tw}$, $w(i)_{yt}$ e $w(i)_{all}$ sono rispettivamente il peso di ogni nodo all'interno degli oggetti sociali della sorgente *Twitter*, *YouTube* e la risultante della combinazione lineare considerando i pesi α e $(1 - \alpha)$.

La figura 5 mostra l'andamento della popolarità dei sei personaggi più commentati su *Twitter* e *YouTube* (con $\alpha = 0.5$). Si può notare come, mentre per la maggior parte dei personaggi l'andamento della popolarità nei vari episodi è abbastanza stabile, quello di Matteo Renzi evidenzia una coppia di picchi in corrispondenza delle puntate in cui è stato ospite della trasmissione. Possiamo inoltre notare come Renzi sia generalmente più popolare su *Twitter* rispetto a Beppe Grillo, che invece risulta più menzionato su *YouTube* (pur non avendo mai preso parte come ospite ad alcuna puntata). La popolarità di Silvio Berlusconi è molto bassa ed infatti egli non figurava, nel periodo di osservazione, tra i possibili candidati per lo schieramento di centro-destra. Si può infine notare come la popolarità di Nichi Vendola assuma un valore apprezzabile solamente nelle settimane che precedono le elezioni primarie della coalizione di centro-sinistra di fine novembre 2012.

Fig. 5 – Andamento della popolarità dei sei personaggi più menzionati all'interno del grafo di conoscenza



4. CONCLUSIONI

In questo articolo è stata descritta l'attività del **Centro Ricerche e Innovazione Tecnologica RAI** nella definizione di un framework per l'integrazione di dati eterogenei provenienti da molteplici sorgenti sia tradizionali, come gli archivi televisivi, le EPG e le ontologie, sia sociali, come *Twitter* e *Facebook*. Il cuore del sistema è rappresentato dalla definizione di una base di conoscenza, basata su un modello a grafo, in grado di rappresentare e integrare in maniera efficiente ed efficace le attività degli utenti. Alcuni esempi hanno permesso di dimostrare come il sistema possa essere utilizzato per implementare innovative metodologie di *analisi inter-sorgente*, facendo, quindi, emergere dai dati l'esistenza di pattern latenti.

Ci si attende quindi che la possibilità di tracciare quella che è stata definita la *seconda vita* del contenuto televisivo potrà influenzare positivamente l'attività di diversi attori. I *broadcaster* avranno a disposizione una nuova tipologia di archivio che permetterà agli esperti del dominio di avere una visione integrata dell'evoluzione nel tempo del contenuto televisivo per quanto riguarda l'interesse degli utenti dei social network, potendo quindi analizzare, tra le altre cose, i trend, le relazioni tra personaggi, l'evoluzione delle opinioni e la popolarità. A livello *editoriale*, la piattaforma di analisi permette di individuare contenuti, personaggi e temi più emergenti e discussi dagli utenti, fornendo quindi un supporto decisionale nella pianificazione del palinsesto e nella creazione di nuovi contenuti. L'*utente finale* potrà godere in maniera indiretta dei vantaggi dell'uso della piattaforma da parte delle media company. Il framework proposto potrà, infatti, essere utilizzato come base per la creazione di *sistemi di raccomandazione* in servizi di *video on demand* e come sorgente di informazioni complementari in *applicazioni second screen*.

L'attività attuale si sta concentrando principalmente su tre fronti: da un lato si intende aumentare il potere espressivo del modello, permettendo quindi di catturare in maniera ancora più accurata gli aspetti rilevanti dell'attività degli utenti sui social network.

Dall'altro lato, l'accuratezza degli strumenti di riconoscimento delle entità e estrazione delle emozioni è un fattore determinante di tutto il sistema e per questo motivo è in corso un ulteriore sforzo verso un continuo miglioramento delle prestazioni. Inoltre, è in fase di realizzazione un prototipo software che rappresenti, per l'utente finale, un sistema di alto livello per la valutazione della popolarità, della reputazione, dei sentimenti e più in generale dei trend dei concetti televisivi.

BIBLIOGRAFIA

- [1] S. Baccianella, A. Esuli e F. Sebastiani, [Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining](#), in "Proc. Of LREC 2010, 17-23 maggio 2010, Valletta, Malta", pp 2200-2204, 2010.
- [2] B. Bara, [Cognitive Pragmatics: The Mental Processes of Communication](#), MIT Press, 2010.
- [3] P. Cesar, D. C. A. Bulterman e A. J. Jansen, [Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content](#), in "Changing Television Environments - 6th European Conference, EUROITV 2008, Salzburg, Austria, July 3-4, 2008 Proceedings", Lecture Notes in Computer Science volume 5066, pp 168-177, Springer, 2008.
- [4] Cypher query language, <http://neo4j.com/developer/cypher-query-language/> (ultimo accesso 15 maggio 2015)
- [5] M. Doughty, D. Rowland e S. Lawson, [Co-viewing live tv with digital backchannel streams](#), in "Proceedings of 9th European Conference on Interactive TV and Video, EuroITV'11, Lisbon, Portugal, June 29-July 1, 2011", pp 141-144, ACM, 2011
- [6] M. Ferraris, [Documentality or why nothing social exists beyond the text](#), in "Cultures. Conflict - Analysis - Dialogue, Proc. of 29th International Ludwig Wittgenstein-Symposium, Kirchberg, Austria, August 6-12, 2006", pp 385-401. Austrian Ludwig Wittgenstein Society, 2006.

- [7] L. C. Freeman, *A set of measures of centrality based on betweenness*, in "Sociometry", Volume 40, Numero 1, pp. 35–41, 1977.
- [8] P. N. Johnson-Laird, *Mental Models*, Cambridge University Press, 1983.
- [9] H. Kopcke e E. Rahm, *Frameworks for entity matching: A comparison*, in "Data & Knowledge Engineering", Volume 69, Numero 2, pp.197 – 210, Elsevier, febbraio 2010
- [10] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, The University of Chicago Press, 1987
- [11] Neo4j Homepage, <http://neo4j.com/> (ultimo accesso 18 maggio 2015)
- [12] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, 2010.
- [13] L. Padró e E. Stanilovsky, *Freeling 3.0: Towards wider multilinguality*, in "Proceedings of LREC 2012, Istanbul, Turkey, May 23-25, 2012", pp 2473–2479, 2012.
- [14] E. Pianta, L. Bentivogli e C. Girardi, *MultiWordNet: developing an aligned multilingual database*, in 'Proceedings of the 1st International WordNet Conference, January 21-25, 2002, Mysore, India', pp. 293-302, 2002
- [15] J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, 1969.
- [16] J. R. Searle, *The Construction of Social Reality*, Simon and Schuster, 1995
- [17] R. E. Shaw e J. E. Bransford, *Perceiving, acting, and knowing: Toward an ecological psychology*, Lawrence Erlbaum Associates, 1977.
- [18] C. Strapparava e A. Valitutti, *Wordnet affect: an affective extension of wordnet*, in "Proceedings of LREC2004, Lisbon, Portugal, May 26-28, 2004", pp. 1083-1086, 2004
- [19] I. S. Dhillon, S. Mallela e D. S. Modha, *Information-theoretic co-clustering*, in "Proceedings of ACM SIGKDD-2003, Washington, DC, USA, August 24-27, 2003", pp. 89-98
- [20] A. Antonini, L. Vignaroli, C. Schifanella, R. G. Pensa, e M. L. Sapino, *MeSoOnTV: A media and social-driven ontology-based tv knowledge management system*, in "Proceedings of the 24th ACM Conference on Hypertext and Social Media, Paris, France, May 1-3, 2013", pp. 208-213, 2013