



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

Social Mining: osservare il comportamento dei consumatori attraverso twitter

Candidato:
Simone Ferrucci

Relatore:
*Prof. Francesco
Varanini*

Anno Accademico 2011-2012

*Ai miei amici e
colleghi.*

Indice

| | |
|---|------------|
| Abstract | 5 |
| Introduzione | 6 |
| 1.Una piattaforma di raccolta dati | 8 |
| La raccolta dei dati..... | 9 |
| Seguire le persone giuste..... | 12 |
| Usare le parole giuste | 13 |
| Un po' del tutto..... | 15 |
| 2.Business intelligence..... | 16 |
| 3.Information Retrieval | 19 |
| 4.Sistema informativo direzionale..... | 21 |
| 5.Sentiment analysis | 22 |
| 6.Analisi delle reti sociali | 24 |
| 7.La mia attrezzatura..... | 27 |
| Un po' di numeri | 28 |
| Hashtag..... | 29 |
| Retweet..... | 30 |
| Grafi | 31 |
| 8.La rete delle <i>hashtag</i> per utente..... | 32 |
| Aprile..... | 33 |
| Maggio..... | 40 |
| Giugno | 46 |
| Luglio..... | 52 |
| Totale | 58 |
| 9.Classificare gli <i>Hashtag</i>: un approccio al <i>sentiment analysis</i> basato sul grafo..... | 64 |
| 10.Rete delle Co occorrenze delle <i>Hashtag</i> | 68 |
| Aprile..... | 68 |
| Maggio..... | 75 |
| Giugno | 81 |
| Luglio..... | 87 |
| Totale | 92 |
| 11.Cosa si dice su Twitter a Giugno..... | 97 |
| 12.Retweet a Luglio | 100 |
| 13.Proteggere i propri dati in un sistema per condividere i propri dati | 102 |
| Facebook e la sua politica..... | 102 |
| Come Facebook spiega la sua politica | 103 |
| Informazioni ricevute da Facebook sugli utenti | 103 |
| Informazioni pubbliche | 106 |
| Utilizzo da parte di Facebook delle informazioni ricevute | 107 |
| Un occhio critico a Facebook..... | 108 |
| Twitter e la sua politica..... | 109 |
| Un occhio critico tra Facebook e Twitter | 111 |

| | |
|------------------------------------|------------|
| 14.Una storia sociale | 112 |
| Conclusione | 115 |
| Glossario | 117 |
| Su Twitter | 117 |
| Analisi delle reti | 117 |
| Bibliografia..... | 118 |

Abstract

Privacy, dato sociale e analisi, sono questi i temi trattati in questo elaborato.

L'obiettivo è quello di sviluppare e testare dei metodi per l'analisi dell'informazione sociale estratta da un socialnetwork.

Nella prima parte del lavoro si riassumono i principali settori scientifici usati come punto di partenza per lo sviluppo del progetto di analisi e raccolta del dato sociale.

Nella seconda parte si descrive l'attrezzatura sviluppata (*script Python*) per raccogliere e analizzare questi dati.

Nella terza parte si esegue l'analisi dei dati estratti, con gli strumenti del passaggio precedente, attraverso delle misurazioni utilizzate nel campo di studio "analisi delle reti".

In fine il lavoro si conclude con una comparazione tra le politiche di privacy tra due *socialnetwork*.

Quest'ultima parte cerca di spiegare perché sia stata analizzata l'informazione contenuta in Twitter e non quella archiviata da Facebook.

Il dato estratto da Twitter è basato su una serie di parole chiave scelte per rappresentare il settore di interesse per la compagnia Autogrill Spa.

Una delle principali domande che quest'azienda si pone è quale sia la sua reputazione sulla rete e a quali concetti vengono associati i suoi marchi e di conseguenza i loro prodotti.

Per rispondere a questo sono state analizzate le reti di retweet e quelle di relazione tra hashtag.

Le prime servono ad analizzare quali siano gli utenti più influenti nelle reti estratte e a capire quanto si diffondono le loro opinioni.

Il secondo tipo di rete serve invece a mostrare quali siano i concetti (hashtag) più vicini ai marchi aziendali e quali sono i più centrali nella rete.

Introduzione

Lo scopo di questo elaborato è **raccogliere ed elaborare dati sociali così da permettere l'individuazione di eventi sociali rilevanti e/o opinioni condivise**. Grazie al Web il dato è in parte disponibile a tutti e necessita solo di essere elaborato nella maniera corretta. I *social network* sono degli aggregatori di questo tipo di dato e in alcuni casi stimolano anche la sua produzione. Per dato sociale s'intende quello in grado di parlarci delle interazioni di un individuo con gli altri o del modo con cui condivide con loro le informazioni in suo possesso.

L'accesso ai dati è regolamentato dal servizio o dall'azienda che li ha raccolti passando attraverso l'accettazione delle condizioni da parte dell'utente. Questa è quella che si definisce politica di *privacy*. Non tutti i servizi hanno la stessa idea di *privacy*, in particolare in questo elaborato vengono comparate le politiche di Facebook e Twitter. **Comprendere le differenze, tra le regole sulla *privacy* dei due social network, è importante per capire se l'informazione sull'utente è proprietà dell'utente stesso, dell'azienda o del mondo. Nell'ultima parte della tesi viene dato spazio a al confronto, tra le politiche di *privacy* dei due Social Network, il quale è uno dei punti centrali dell'elaborato.**

In letteratura si trovano molti lavori scientifici che trattano il come raccogliere dati sociali e come analizzarli, ma questo settore è in rapido sviluppo e le sue tecniche cambiano continuamente. **Non esistono strumenti già pronti per il tipo di analisi che è stata portata avanti, per questo lavoro sono stati sviluppati dei programmi appositi che gestiscono le varie fasi del processo e che si trovano ancora in una fase di sviluppo.**

Lo scopo finale di questo elaborato è di aiutare la comprensione di cosa dica la società del web di un determinato argomento. Gli strumenti utilizzati hanno permesso di raccogliere molta informazione in grado di essere di aiuto a chi vuole comprendere. Questo lavoro è solo la punta di un immenso

Iceberg che è il *Social Data Mining*¹. Tale settore è una delle attuali frontiere degli studi sociali, economici, accademici e informativi.

Per contestualizzare il lavoro sono stati descritti i settori scientifici da cui questo tipo di lavoro attinge come *Information retrieval*, *Business intelligence*, Analisi delle reti e *Opinion mining*; tali argomenti verranno affrontati nei primi capitoli. Il capitolo finale affronta il problema della *privacy* proponendo un'analisi delle politiche di Facebook e Twitter.

¹ “Datamining nel social web” Matthew A. Russel, O’Reilly

1. Una piattaforma di raccolta dati

L'informazione è una risorsa utilissima, può portare a una migliore comprensione degli eventi o all'identificarne di nuovi. Se ben utilizzata e al momento giusto può trasformarsi in conoscenza e da lì portare un contributo essenziale al suo possessore.

Spesso però l'informazione non è da subito disponibile, si presenta sotto forma di dato. Detto questo il dato assume un ruolo decisamente importante nel processo di creazione della conoscenza, quindi è logico partire da esso, bisogna solo decidere quale tipo di dato andare a cercare e dove. La scelta dipende dal tipo d'informazione che ci serve, ma soprattutto dal tipo di conoscenza che speriamo di estrarne.

Lo sviluppo delle tecnologie *ICT* ha permesso la generazione di una sconfinata quantità di dati sparsi pe tutta la rete, sconfinata è decisamente il termine adatto visto che non si conosce la dimensione che il web ha raggiunto. Sul web possiamo trovare informazioni di borsa, notizie, recensioni, racconti, voci di un'enciclopedia (non necessariamente Wikipedia).

Uno degli obiettivi di questo elaborato è la teorizzazione di una piattaforma di raccolta di dati provenienti dai *social network*.

Queste piattaforme giocano un ruolo cruciale nel processo di creazione dell'informazione sulla rete poiché mettono a disposizione, dietro le giuste protezioni, i dati privati degli utenti, i quali decidono di condividerli con i loro amici o in qualche caso con il mondo intero, gli approfondimenti sono lasciati al capitolo dedicato alla *privacy* sui principali *social network*.

Utilizzando le *Api*² dei vari *social network* è possibile avere accesso a vari tipi di informazione come:

- Dati anagrafici
- Lavoro
- Interessi
- Amici
- Familiari

² http://it.wikipedia.org/wiki/Application_programming_interface

- Opinioni(Sotto forma di post o Tweet)

In questo elaborato verrà preso in esame, come sorgente di dati, Twitter. La sua politica sulla *privacy* e il tipo di dato che gli utenti condividono lo rendono il candidato ideale per molti lavori di ricerca scientifica.

La raccolta dei dati

Abbiamo già detto che il dato è importante e che il web ne mette a disposizione una grande quantità quindi passiamo a vedere come se ne può ottenere una parte. La raccolta di dati può avvenire con diverse metodologie, a seconda della piattaforma che stiamo interrogando, del tipo di dato e della sua struttura. In alcuni casi il dato non ha una struttura data a priori e in questo caso si parla di dato non strutturato. L'ipotesi qui proposta è questa:

« Costruire una macchina che raccolga il dato prodotto da Twitter. Questo dato sarà poi interrogabile per rispondere a delle domande poste dall'utente. L'utente potrebbe non sapere a priori cosa sta cercando di preciso. »

L'utente potrebbe non sapere cosa cerca, potrebbe solo voler sapere cosa si dice di una certa cosa su Twitter. Partirà con dei tentativi generici e, se esiste, scoprirà la chiave per quello che cerca, questo processo si definisce *Abduzione*³.

Il dato principale su Twitter sono i messaggi degli utenti (*Tweet*); esistono diverse strategie per raccogliere questo dato, ma a seconda di quale si utilizza il dato può avere delle caratteristiche di un certo tipo o di un altro. In letteratura si trovano principalmente tre metodologie di approvvigionamento:

³ http://en.wikipedia.org/wiki/Abductive_reasoning

- Seguire un utente;
- Utilizzare delle *keyword*;
- Collegarsi al flusso dei *tweet*.

Il primo metodo prevede di scegliere un gruppo di utenti secondo l'utente rilevanti per l'informazione che cerca. Questo sistema è molto diffuso tra chi cerca un'informazione proveniente da un solo ambito. Prendendo per esempio la politica; se servono informazioni su cosa i politici dicono online basta scegliere un campione di essi abbastanza omogeneo e seguirlo.

Questa strategia di raccolta però viola uno dei punti fondamentali della macchina da costruire: "L'utente potrebbe non sapere cosa sta cercando".

Il secondo metodo consiste nello scegliere delle parole chiave adatte a quello l'utente va cercando o pensa di cercare. Twitter mette a disposizione l'*Api Search*, che prende in ingresso la parola chiave e altri attributi (es: nazione, codifica, numero di risultati per pagina). Per ogni *Keyword* possiamo ottenere massimo 10 pagine di risultati ognuna delle quali può contenere al massimo 100 risultati il che fa mille risultati massimi a richiesta. Bisogna tener presente che Twitter rende disponibile i messaggi fino a 6 giorni a dietro da quando è stata posta la richiesta. Il numero di interrogazioni è limitato da una quota oraria di 500 richieste.

La terza soluzione permette di ottenere una percentuale dei *tweet* di Twitter, è possibile anche specificare dei limiti di nazione. Per gli utenti comuni la percentuale di *tweet* scaricabile è dell'1%, per scopi scientifici o dietro pagamento è possibile avere un'*Api* speciale che permette di scaricare il 10% del traffico.

I tempi di attivazione per scopi scientifici di questa "Super *Api*" sono molto lunghi, in alcuni casi si parla anche di otto mesi, mentre non ci sono dati disponibili per sapere quanto tempo ci voglia dietro pagamento è noto solo che ci sono società esterne a Twitter le quali mettono a disposizione questo servizio dietro pagamento.

La tabella seguente cerca di riassumere pregi e difetti delle tre soluzioni

| | Seguire gli utenti | <i>Keyword</i> | <i>Stream</i> |
|--------|---|---|--|
| Pro | <ul style="list-style-type: none"> - Semplice - Il dato raccolto ha un'alta probabilità di essere attinente | <ul style="list-style-type: none"> - Simile ad una ricerca su internet - Permette di realizzare delle interrogazioni su più domini - Può essere espansa con facilità (Aggiungere <i>keyword</i>) | <ul style="list-style-type: none"> - Eterogenea, si ottengono dati da diversi domini d'interesse - Il dato è del tutto imprevisto |
| Contro | <ul style="list-style-type: none"> - Limiti molto forti dati a priori (Dobbiamo conoscere chi è interessante) | <ul style="list-style-type: none"> - Senza le <i>Keyword</i> giuste i risultati non sono utili - Parte del dato estratto potrebbe non essere rilevante per la ricerca (rumore) | <ul style="list-style-type: none"> - Dati troppo sparsi - Il campione può non essere rappresentativo della realtà nel caso dell'1% |

Il fatto che Twitter metta a disposizione i *tweet* solo fino a sei giorni a dietro crea il problema delle scelte, se nella prima soluzione l'utente sceglie le persone sbagliate da seguire non troverà l'informazione che cerca e dopo mesi di analisi non può tornare in dietro e cambiare il campione. Nella seconda soluzione il problema è simile, se le parole chiave sono "sbagliate" non si può tornare in dietro per cambiarle. Il problema diventa quindi quello di avere la sorgente giusta.

Seguire le persone giuste

Immaginiamo il problema di rispondere alla domanda «Quali sono gli argomenti principali trattati dai vari politici?». Per rispondere a questa domanda si potrebbero seguire gli account dei principali politici d'interesse nazionale, Aggregare tutti i loro rispettivi messaggi e cercare di estrarre le parole più rilevanti. Per la raccolta dei dati questa potrebbe essere un'ottima strategia. Ma se la domanda è cosa «Cosa si dice del politico X?», In questo caso è difficile mettere insieme una strategia di raccolta dati. Si potrebbe ipotizzare di seguire i personaggi più autorevoli in materia, ma questi non parleranno esclusivamente del politico di nostro interesse e allora si genererà del rumore che può falsare le analisi e portarci a delle conclusioni sbagliate.

Come si fa poi a dire chi è autorevole e chi no, si potrebbe utilizzare la conoscenza personale dell'individuo, ma questa potrebbe essere un grandissimo ostacolo sulla strada per trovare la giusta risposta alla domanda. Spesso su Twitter si cerca di attribuire l'autorevolezza in base al numero di *Follower* che un utente ha, ma questo è un metodo sbagliato per due ragioni:

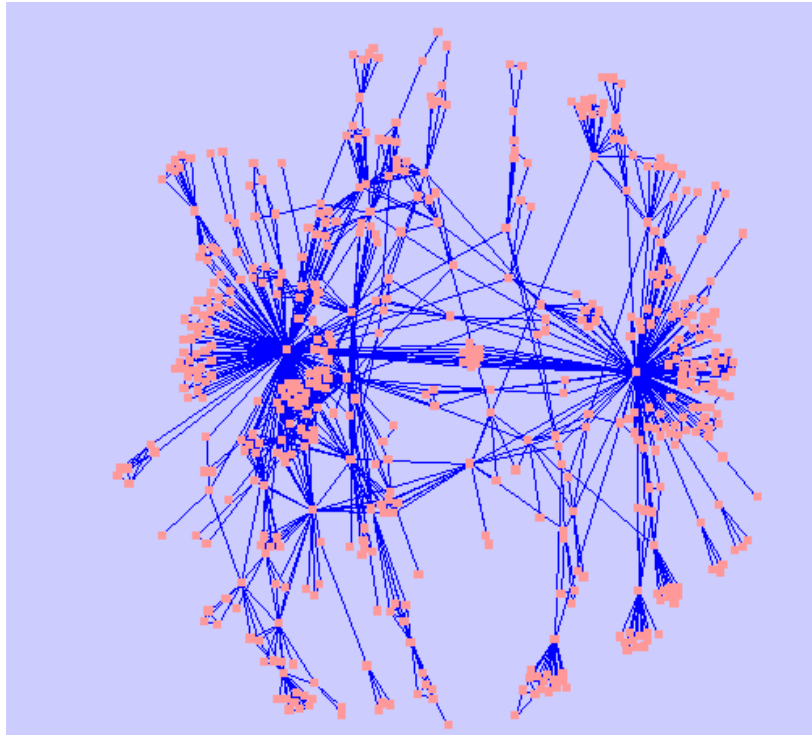
1. L'utente potrebbe parlare di più ambiti completamente differenti;
2. L'utente potrebbe aver acquistato quelle amicizie.

Il punto due è dovuto al fatto che sono nate delle aziende in grado di fornirti un alto numero di *follower* dietro pagamento. Il corriere della sera ha pubblicato un articolo⁴ su di un imprenditore che afferma di aver comprato 50 mila *follower*.

Per comprendere chi sia autorevole o meno servono altri tipi di misurazione, come per esempio il numero di *retweet*, oppure delle misurazioni basate sul grafo. Per *retweet* s'intende quando una persona diffonde il messaggio scritto da un altro utente. In Questo elaborato

⁴http://www.corriere.it/scienze_e_tecnologie/12_maggio_16/compravendita-follower-fan-facebook-twitter_1d5ee5ec-9f8e-11e1-b258-f2fcbb76be58.shtml

s'ipotizza che ricostruendo il grafo di chi a twittato e da chi è stato retwittato può essere uno dei metodi per costruire una rete di relazioni reali su Twitter. Successivamente su questa rete si possono fare delle misurazioni come la centralità di un nodo, il suo grado, la sua influenza nella rete, ecc...



Lo studio dell'influenza sulla rete è rimandato al capitolo apposito sulle reti sociali e sugli effetti a cascata.

Usare le parole giuste

Volendo rispondere alla domanda «In che contesto è usato il termine *Lampredotto?*» si potrebbe partire dal chiedere a Twitter tutti i più recenti messaggi, disponibili, nei quali ricorre la parola *lampredotto*. A questo punto andrebbero analizzati i dati raccolti per capire se abbiamo utilizzato la parola chiave giusta e se ce ne sono altre che potrebbero espandere la ricerca o indirizzarla meglio. Molto probabilmente, i *tweet* ottenuti, o sono troppo pochi o non si distribuiscono su di un arco temporale abbastanza ampio da darci l'idea di come realmente si usa il termine es:

- Viene fatto un massimo di 10 *tweet* la settimana con la parola chiave ricercata, in questo caso avremmo ottenuto un massimo di 10 messaggi.

- Negli ultimi due giorni sono stati fatti più di 1000 messaggi con la parola scelta, in questo caso noi otterremmo solo questi e il campione sarebbe limitato nel tempo.

Sarebbe utile quindi seguire lo sviluppo dell'uso del termine, o dei termini, per un periodo abbastanza lungo come un mese o più.

Anche questo sistema però può risentire di scelte iniziali sbagliate o fin troppo soggettive, portando al non avere il dato per rispondere alla domanda.

Un Hashtag adeguata

Per risolvere il problema sopra esposto si potrebbe utilizzare come *Keyword* delle *hashtag* selezionate tra quelle utilizzate dagli utenti.

Per hashtag s'intende una parola preceduta dal simbolo "#", questa è una caratteristica che secondo gli utenti in aumenta l'espressività di Twitter.

Supponiamo di aver scaricato un tot di messaggi con la tecnica *Keyword*, se il campione di messaggi è rappresentativo si potrebbe estrarre gli *hashtag* contenuti nei messaggi e creare una rete delle loro relazioni. Se due hashtag appaiono nel solito messaggio esiste un *link* tra di loro.

Da questa rete si estrarranno le *hashtag* più rilevanti che poi verranno utilizzate come *keyword* per la ricerca. I problemi però sono due:

- L'insieme degli *hashtag* andrebbe aggiornato costantemente;
- Su quali basi si sceglie un *hashtag* o un altro;

Un po' del tutto

Utilizzare l'Api che permette di scaricare una percentuale del totale flusso dei messaggi potrebbe tornare utile a chi deve realizzare un'analisi linguistica ad ampio spettro, senza porsi il problema di scegliere dove prendere il dato.

Per la macchina richiesta invece questa soluzione non sarebbe ottimale a meno che non si riesca a scaricare una percentuale molto significativa del flusso; il 10% sarebbe un buon inizio, ma potrebbe anche non contenere niente di quello che stiamo cercando. Altro problema è capire se questa percentuale del flusso sia rappresentativa o meno:

- È un campione accettabile?
- Con quali metodologie è stato scelto?

Sarebbe interessante riuscire a parallelizzare più macchine ognuna delle quali scarica il 10% dei messaggi da una zona specifica e poi mettere insieme il tutto. In caso contrario l'1% di tutto il mondo rischia di essere un campione troppo sparso, e forse lo sarebbe anche il 10%. Va ribadito il fatto che non è una questione di dimensioni, ma di qualità (Attinenza).

2. Business intelligence

Uno dei principali scopi dell'informazione sociale è la *Business intelligence*.

*«I sistemi di Business intelligence (BI) offrono supporto al processo di pianificazione e decisione rendendo disponibili informazioni sul contesto competitivo di un'organizzazione. **Le caratteristiche stesse dell'analisi competitiva non permettono di conoscere a priori quali dati saranno necessari per ricavare le informazioni utili al processo decisionale.** (...) un sistema di Business intelligence è volto al supporto di decisioni manageriali e strategiche di carattere competitivo effettuate dalla direzione aziendale. Ne consegue che i dati analizzati non saranno prevedibili a priori e che le informazioni trattate potrebbero provenire da **fonti differenti e non omogenee**, siano esse interne o esterne all'organizzazione. (...) le fonti esterne vengono utilizzate per completare o migliorare quelle interne (...)»*

Sistemi per la gestione dell'informazione

Questa definizione spiega cosa un sistema di *Business Intelligence* dovrebbe essere, nelle grandi linee, ma forse per comprendere meglio cosa c'è dietro è necessario partire dagli albori della materia.

Fare *business intelligence* vuol dire dare le giuste informazioni per permettere di rispondere alle domande di chi fa le scelte di mercato.

Le scelte intelligenti possono essere fatte solo con la giusta conoscenza. Chiunque ogni giorno deve fare numerose scelte ed è guidato in esse dalle informazioni che ha, se ha poche informazioni i rischi di una scelta aumentano, al contrario più informazione si ha e più il rischio diminuisce. I dati a disposizione di questi sistemi riguardano principalmente i prodotti dell'azienda, i suoi clienti, i processi produttivi e l'azienda stessa. Immaginatoci che un direttore di una catena di negozi deva premiare il più produttivo negli ultimi tre mesi in base al rapporto tra fatturato e numero di scontrini prodotto, la piattaforma deve essere in grado di fornire l'informazione per fare la scelta giusta.

I dati sono una risorsa molto preziosa che deve essere conservata in un modo adeguato.

Prima degli anni '60 i dati erano contenuti in archivi cartacei, successivamente con l'avvento dei computer alcune aziende hanno iniziato ad utilizzare il supporto digitale. Per garantire una maggiore sicurezza dei dati, visto la durata media dei vecchi supporti digitali come il *floppy*, la soluzione più logica era quella di utilizzare un Database, il quale fu inventato da Edgard Codd nel 1969⁵. Successivamente all'inizio degli anni '70 apparirono le prime applicazioni software che permettevano di utilizzare il *database* anche a chi non era un esperto informatico. Grazie alla facilità di utilizzo la mole di dati disponibile aumentò considerevolmente. Il problema a questo punto era di immagazzinare diversi tipi di dato provenienti da diverse tipologie di sistemi e per questo iniziò un ampio utilizzo di *Data Warehouse* i quali aggregano informazione proveniente da diverse sorgenti e permettono ricerche su più dimensioni. Il *Data Warehouse* è diventato popolare all'inizio degli anni '80, e due dei suoi padri sono Ralph Kimball⁶, il quale è un autore di libri sulla *business intelligence* e sul *data warehousing*, e Bill Inmon. Con questa tecnologia i dati possono essere maneggiati facilmente anche se frammentati o provenienti da diverse sorgenti.

Nel 1989 nasce il termine *Business Intelligence*, utilizzato da Howard Dresner per descrivere «concetti e metodi per migliorare le decisioni lavorative usando un *software* di supporto basato sui dati».

Con gli anni il numero di dati disponibili è cresciuto moltissimo fino ad esplodere con l'avvento di internet e delle tecnologie informatiche che riempiono la vita dell'individuo moderno. Uno dei problemi attuali con tutta questa mole d'informazioni è l'accesso. Spesso l'informazione che si cerca non è immediatamente disponibile perché frammentata in varie parti o perché da elaborare. I *software* di gestione spesso non sono così intuitivi. Bisogna quindi contattare l'esperto di informatica del settore di informatica del settore il quale comporrà i dati nella maniera giusta per rispondere alla domanda. Ma spesso questo rallenta molto il processo, e le risposte alle

⁵ http://en.wikipedia.org/wiki/Edgar_F._Codd

⁶ http://en.wikipedia.org/wiki/Ralph_Kimball

domande invece devono essere veloci. Il sistema di BI dovrebbe permettere alle persone di maneggiare facilmente i dati e di avere dei riassunti semplici della situazione attuale senza dover modificare il funzionamento del sistema. Riassumendo, più il sistema è intuitivo più è utile ed efficiente.

Uno dei problemi che occorrono con i nuovi tipi di dato, provenienti principalmente dal web, è che questi non sono strutturati, quindi la loro analisi richiede tecnologie differenti. Di questo si occupa l'*Information Retrieval*.

3.Information Retrieval

Recupero d'informazioni può voler dire veramente molte cose; pensando di voler acquistare una casa significa trovare delle informazioni con cui scegliere, preparando una relazione accademica vuol dire trovare le fonti giuste per un'analisi approfondita e corretta dell'argomento. In informatica descrive quella disciplina che si occupa del recupero d'informazione da una sorgente non strutturata o semi strutturata.

«*Information Retrieval* (IR) si occupa di trovare materiale (di solito documenti) di natura non strutturata (testo) che soddisfa alcuni bisogni d'informazione ricercandolo in una grande collezione di elementi»

Introduction to Information Retrieval.

Materiale non strutturato significa che non ha una struttura semantica data a priori e quindi di facile consultazione per il computer. Per esempio il libero testo, anche se per l'essere umano la struttura del testo (paragrafi, titoli, ecc..) è ovvia questo non vuol dire che lo sia per il computer; il quale se non programmato a dovere non sa ne come dividere le parole ne come riconoscere l'inizio di una frase o la fine di un'altra. Tra informazione strutturata e informazione non strutturata vi è l'informazione semi-strutturata per esempio un documento codificato in xml.

Obiettivo dell'IR è anche il supporto all'utente nel capire quanto un dato documento possa essere rilevante per la ricerca effettuata, prendiamo in esempio il *ranking*. Anche il raggruppamento di documenti simili rientra nelle sue applicazioni, organizzare le risorse in base al contenuto è un po' come organizzare i libri in base all'argomento o alla tipologia. La classificazione è un'altra applicazione: dati un insieme di argomenti decidere in quale di questi rientra il libro X.

Negli ultimi anni è diventato molto importante il recupero d'informazioni personali. Pensiamo ai sistemi operativi che integrano metodologie (strumenti di ricerca) per aiutarci nel trovare documenti nel nostro computer, oppure al motore di ricerca del servizio di email, alcuni di questi sono in grado di capire le nostre tipologie d'interessi e di classificare la

posta in base a questi o meglio ancora a distinguere ciò che è *spam* da ciò che non lo è.

Anche i motori di ricerca come Google o Bing rientrano nel settore dell'IR, interrogando grandi quantità d'informazione semi-strutturata (html) ci aiutano a scegliere quale di questa fa al caso nostro, spesso in maniera molto sofisticata, basta pensare agli algoritmi di *ranking* e a quelli basti sull'affidabilità delle pagine.

4.Sistema informativo direzionale

Un sistema informativo raccoglie l'insieme di dati prodotti dall'azienda e li trasforma in informazione utile a essa. Ogni processo produttivo produce dell'informazione, basta pensare a una catena di negozi e al loro sistema di gestione della cassa. Ogni singola transazione viene archiviata in un database, per far funzionare il software di gestione.

Questi dati possono essere estratti dal database ed estrapolati da loro contesto per diventare nuova informazione. I *Data warehouse* immagazzinano questa informazione e la rendono facilmente disponibile per l'analisi. Supponendo che un impiegato della sezione *marketing* voglia vedere l'andamento di una tale categoria di prodotti per un dato periodo gli basterà interrogare il *data warehouse*, anche senza aver progettato in precedenza di raccogliere questo dato, visto che esisteva già come dato nel database delle casse. Il *data warehouse* fornisce anche i *tool* di analisi per trattare i dati da esso raccolti. L'obiettivo principale di un *data warehouse* è principalmente di fornire l'informazione giusta a coloro che devono prendere delle decisioni.

Un sistema che permette a chi dirige di prendere delle decisioni basate sui dati è detto **Sistema Informativo Direzionale**.

Nei *data warehouse* l'informazione è strutturata su più dimensioni, anche al di là delle sole 3 intuitive. L'analisi multidimensionale dei dati è chiamata OLAP: *On Line Analytical Processing*.

L'aggregazione di informazioni per una data area dell'azienda si dice **Data Mart**.

5.Sentiment analysis

Con questo termine ci si riferisce a un filone di studi il cui obiettivo è quello di estrarre informazioni utili per comprendere l'opinione di un gruppo di soggetti su un argomento, questo settore di studi è fortemente correlato con l'*information retrieval*. In alcuni casi si cercano messaggi che parlino di un argomento sotto analisi, in altri si cerca di capire se un testo è positivo o negativo rispetto all'argomento e in altri ancora si cerca di fare le due cose insieme.

Il *sentiment analysis* è spesso collegato con l'*information retrieval* in quanto il primo può essere visto come una sottoclasse del secondo, il quale come già visto si occupa di fornire i materiali più attinenti per rispondere alla domanda che gli viene posta.

Uno dei principali obiettivi del settore attualmente è quello di dare una polarità alle singole parole, se sono positive o negative e quanto lo sono.

Si possono distinguere due principali tipologie di approccio alla materia:

- *Approccio linguistic-based*, si cerca di definire uno *score* di opinione politica tramite algoritmi di *machine learning* che analizzano le porzioni di testo prodotte dagli utenti;
- *Approccio graph-based*, si cerca di definire uno *score* di opinione politica analizzando la rete di contatti dell'utente (*homophily*: tendenza degli individui ad associarsi e a stringere legami con altri individui a loro simili sotto qualche aspetto).
-

In alcuni casi il *sentiment analysis* non si limita a dare una polarità alle parole o ai testi, ma bensì agli utenti.

Una volta identificati dei possibili utenti negativi, o positivi, si può generalizzare la scoperta presumendo che individui con interessi simili o amicizie simili siano più portati a condividere la propria opinione che con persone completamente differenti. La ricerca in base all'omofilia ha ricevuto negli ultimi anni molta attenzione e non solo nel recupero d'informazione, ma anche nell'analisi delle reti.

Twitter è stato identificato come una delle principali sorgenti d'informazione per questo tipo di studi, e ogni anno fioriscono teorie e strumenti per migliorarne l'analisi.

Prendiamo per esempio i ricercatori di *TweetFeel* i quali cercano di realizzare una veloce analisi in base ad una *Keyword* che l'utente gli passa. Scarica tutti i *tweet* recenti per quella parola chiave e poi utilizza un classificatore automatico per cercare di distinguere la polarità.

In alternativa ai progetti già strutturati esistono basi di conoscenza interrogabili per comprendere la polarità di una parola, un esempio è *SentiWordNet* dove le parole sono classificate da otto classificatori differenti, purtroppo non ci sono molte risorse di questo tipo per la lingua italiana.

Uno delle sfide che questo settore sta cercando di vincere è il riconoscimento del sarcasmo il quale si basa sull'utilizzo di parole con polarità opposta a quello che veramente si comunica.

6. Analisi delle reti sociali

L'analisi delle reti è un settore di ricerca che ha visto negli ultimi anni aumentare esponenzialmente la quantità di dati da analizzare.

Secondo molti studiosi universitari, tra cui Barabasi⁷, la rete è lo strumento ideale per analizzare i fenomeni complessi.

Un evento complesso è costituito da più oggetti connessi tra di loro. Per comprendere questa tipologia di evento bisogna studiare le relazioni tra gli oggetti del sistema.

Una rete può essere lo strumento giusto per rappresentare diversi tipi di relazioni. Pensiamo a:

- Parentele;
- Organigramma Aziendale;
- Amicizie;
- Interessi;
- Opinioni;
- Luoghi visitati.

La prima grande rete è stata internet, che ha fornito una grandissima quantità d'informazione con la propria struttura sviluppata in maniera "selvaggia". Nessuno ha detto quale forma internet non possa prendere, un cittadino può costruirsi la sua rete personale e collegarla a internet.

Senza lo studio approfondito delle reti interne agli eventi complessi è difficile trovare un modo per spiegarli.

Le reti sociali sono una sottocategoria delle reti: rappresentano relazioni di tipo sociale tra una grande quantità d'individui.

La società è uno dei sistemi più complessi al mondo visto il numero d'individui e il gran numero d'interazioni possibili.

Negli ultimi anni è diventato non solo possibile studiare le reti, grazie ai dati, ma addirittura è diventato necessario comprendere i fenomeni

⁷ <http://barabasilab.neu.edu/courses/phys5116/>

complessi per risolvere problemi della società (per esempio la mobilità) o addirittura prevenirli (Epidemie).

La storia dell'analisi delle reti ha origine con Eulero nel 1735. Eulero è riuscito a gettare le basi della teoria dei grafi, materia su cui si basa l'analisi delle reti.

Nel 1736 Eulero risolse il problema dei sette ponti di *Koninbergh*. La domanda era se fosse possibile con una passeggiata seguire un percorso che attraversi ogni ponte una volta sola e ritorni al punto di partenza.

Eulero dimostrò che non era possibile a causa del numero dispari dei ponti.

Questa soluzione dette origine alla teoria dei grafi che si è poi evoluta nella ben nota topologia.

Negli anni '30 del XX secolo, Jacob Levi Moreno contribuiva a gettare le basi dell'analisi delle reti sociali, la branca della sociologia che si occupa di valutare quantitativamente l'apporto del ruolo di un individuo al gruppo o alla società analizzando la rete di relazioni da cui questa è formata.

Alcune delle sue idee sono state pubblicate nel libro «*who shall survive?*» del 1934 (l'autore è sempre Moreno).

Negli anni '60 poi è nata internet partendo dalla ricerca sul sistema di scambio dei pacchetti.

Nel 1960 Erdosh sviluppa il suo modello dei grafi random, ovvero:

«Un grafo random è composto da N nodi dove ogni coppia di questi è connessa con probabilità P »

Questi modelli si discostano molto da quello che succede nella realtà. La spiegazione più probabile è che esistono molte altre variabili, di cui il modello non tiene conto, che influenzano la rete. Questo modello ha fornito un grande strumento di confronto con le reti reali.

Nel 1967 Stanley Milligram porta avanti uno studio per misurare la distanza minima tra le persone. Il suo lavoro proponeva la tesi che tra tutte le persone esista un grado medio di distanza vicino a 6. Vari studi moderni hanno validato la tesi di Milligram su grandi quantità di dati.

Milligram regala all'analisi delle reti sociali anche un altro concetto fondamentale: la forza dei legami deboli.

Secondo questa teoria, sono i legami deboli, cioè quelli con persone più distanti e che vengono frequentate di rado, a fornire gran parte dello scambio di informazione nella rete. Anche questa teoria è stata validata con studi recenti

Nel 1979 si ha la nascita dello studio delle reti ecologiche intese come ecosistema vivente.

7. La mia attrezzatura

Come si è già visto l'informazione richiede strumenti per essere raccolta immagazzinata e utilizzata; in alcuni casi esistono già strumenti adatti, in altri bisogna riadattare quelli che esistono e in fine casi in cui si rende necessario creare tutta la struttura da zero. Quando è iniziato lo sviluppo del sistema, l'obiettivo era di costruire un *dataset* e analizzarlo.

L'oggetto della ricerca era capire cosa si dice di un'azienda su internet. Visto quanto immensa è la rete e stimando le risorse disponibili, il campo di ricerca è stato ridotto a Twitter. Lo sviluppo del sistema ha richiesto quattro mesi di lavoro durante i quali si è andato a delineare cosa serviva e quindi cosa il sistema dovesse essere in grado di analizzare.

Il punto di partenza è stato il *tweet*, il solo singolo messaggio, scaricato e archiviato utilizzando delle *Keyword*. Successivamente è nata la necessità di analizzare le relazioni di *retweet* tra gli utenti, questo ha permesso di scoprire come si diffondono le opinioni dell'azienda su Twitter; i singoli *tweet* riguardanti l'azienda si considerano come opinioni.

In un secondo momento è emerso l'interrogativo su come gli *hashtag* siano utilizzati e che relazioni ci siano tra di loro, per questo sono state sviluppate specifiche funzioni in grado di estrarre dal messaggio questa informazione.

La stessa cosa è stata fatta per i "Via" e le "referenze":

- Via: Da quale *blog* è stato inviato il messaggio.
- Referenze: Se e a chi era indirizzato il messaggio.

Ho provveduto poi a permettere la segmentazione dei dati per settimane e per mesi sotto consiglio del professor Dino Pedreschi.

Adesso il sistema è in grado di fare le seguenti cose :

- Quanti *tweet* nel tempo;
- Quali *retweet* ha fatto;
- Quali *keyword* ha utilizzato;
- A chi ha indirizzato i *tweet*;
- Quali via ha usato;

- Lista dei *tweet*;
- Lista delle *hashtag*;
- Riferimenti ha persone;
- *Hahtag* nel tempo;
- Stampa di tutti i *tweet* per una keyword nella settimana o nel mese selezionato.

Un po' di numeri

In questa sezione vengono mostrate alcune delle informazioni di base estraibili con gli strumenti sviluppati.

Per prima cosa ecco alcune delle informazioni sui messaggi raccolti partendo da alcune *query* sulle chiavi:

- autogrill;
- muffin;
- spizzico;
- burger king;
- burgheer king (qualcuno sbaglia a scrivere)

La prima statistica da analizzare è “quante parole chiave sono state utilizzate dall'utente x? E quante volte?”.

Il sistema risponde alla domanda restituendo per ogni utente il numero dei suoi *tweet* che contengono la parola.

Esempio:

Greg Turner

```
{'pizza hut': 23, 'starbucks': 1, 'burger king': 21}
```

Questo risultato si legge come “Greg Turner, ha utilizzato il termine Pizza Hut 23 volte, Burger king 21 volte e solo una volta Starbucks.”

Ovviamente per rispondere alla nostra domanda la ricerca va limitata alle parole chiave scelte.

- R U Giulia
{'autogrill':2,'burger king':1}

- Giorgio Pappallo
{'autogrill':1,'burger king':2}

Questa è solo una piccola porzione del risultato di un'interrogazione. Dai risultati si nota che non hanno utilizzato tutte le parole richieste. Il dato che viene estratto può essere utile per capire gli interessi di un utente e vedere se ci sono gruppi di utenti simili.

Hashtag

Possiamo chiedere al sistema di restituire, per ogni utente, gli *hashtag* che ha usato e il numero di volte.

- Tory Jones
{sagittarius:1, libra:,1, scorpio:,1,cancer:1, capricorn:1, gemini:1, pisces:1, virgo:, 1,retweet:1}

- Intervistato.com
{ifj12:1, benetton:1, eni:1, autogrill.:1, asinararevolution'2, ijf12:1, uberlusconi:1}

- Ankon Anconaline
{u'autostrada':1, u'a14':1, ancona:1, viabilità:1, senigallia:1}

- Carlos Campis
{miami:1, airport:1, coffeetime:1, cinnamondolcelatte:1, starbucks:1}

Ovviamente il numero di *hashtag* usati da un individuo medio su Twitter è abbastanza alto, alcuni arrivano anche a 30. Questo tipo d'informazione

può portare alla scoperta di *hashtag* fortemente correlati con dei prodotti o con altre importanti per l'analisi del prodotto ma di cui non sera a conoscenza nel principio.

Retweet

L'analisi dei *retweet* può dire molto di chi come l'informazione si diffonde sulla rete. Il grafico di queste relazioni mostra principalmente la rilevanza dell'utente retwittato e di come le sue idee si diffondono tra gli utenti.

- Giulia Motta
{STELLAMARCIANO:1, FabriziodiPerna:1, ZiaFefe:1, VirginioNero:1, chiara2802:1, ValeVirginauta:1, Chiaia_P:1, SpanuLuca:1, IleniaDeVita:1, ele_instabile:1, SahGermanottaVi:1}
- Noemi Motta
{Deni5e_Motta5:1, ChiaGuerci:1, VincenzoNovari:1, tznferrosite:1, simonesli:1, Lucaargentero:1}

I *tweet* nelle settimane

Un'altra funzionalità interessante per il monitoraggio degli utenti è il numero dei loro messaggi diviso per settimane.

- Alessandro
Aprile [0,2,1,3]
Maggio [0,2,0,3]
- Radio Aldebaran
Aprile [0,0,0,1]
Maggio [2,2,0,4]

Per ogni utente ci sono i vari mesi, rappresentati dalle parentesi quadre, ognuno dei quali si divide in settimane.

Quelli appena presentati sono alcuni metodi basilari per l'analisi del comportamento degli utenti. Ovviamente i dati sono estratti nei mesi di aprile e maggio, utilizzando le parole chiavi scelte.

Grafi

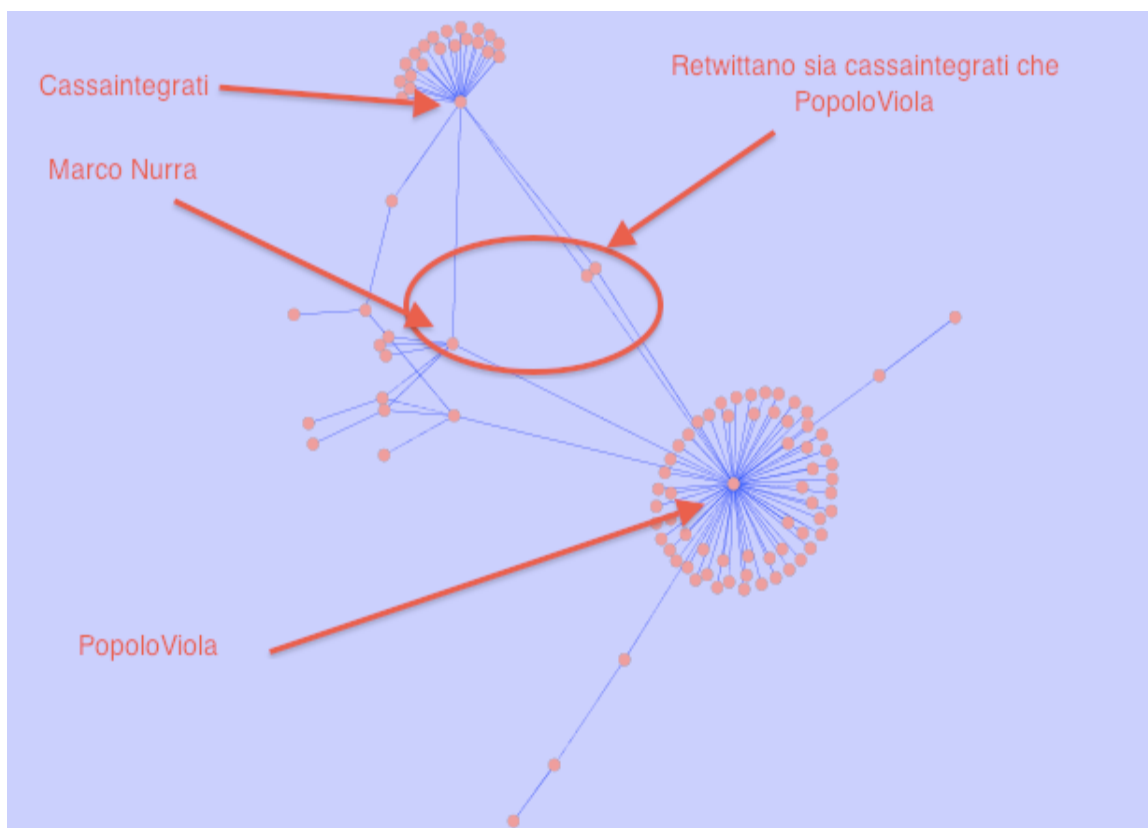
In questa sezione si mostrano alcuni grafi(reti) interessanti. Per prima cosa partiamo dall'analisi dei *retweet*. Ogni persona che ha Twittato è un nodo, esiste un link tra le persone se una delle due ha Retwittato l'altro.

Ho ristretto la ricerca solo ai messaggi per la parola chiave "autogrill" nel mese di aprile.

Nella prima settimana non è accaduto niente di rilevante, nella seconda settimana invece ci sono due nodi con molti Retweet, rispettivamente «PopoloViola» e «Cassaintegrati», ma solo tre utenti hanno retwittato entrambi i nodi e solo uno di questi è stato retwittato a sua volta (Marco Nurra).

I Tweet sono i seguenti:

- PopoloViola: *«76 lavoratrici buttate sulla strada, liquidate anche dopo 20 anni di servizio in Autogrill (dei #Benetton) che NON è in crisi. Gravissimo.»*
- Cassintegrati_: *«La famiglia #Benetton non solo non ritira i licenziamenti #Autogrill ma li raddoppia! 76 lavoratrici tra Roma, Milanoe Bologna.»*



8. La rete delle hashtag per utente

L'*hashtag* è una caratteristica importantissima della comunicazione su Twitter ed è ritenuta, non solo dagli utenti ma anche dagli studiosi⁸, una grande fonte d'informazione per capire l'argomento del messaggio e in alcuni casi anche il giudizio dell'utente sull'argomento.

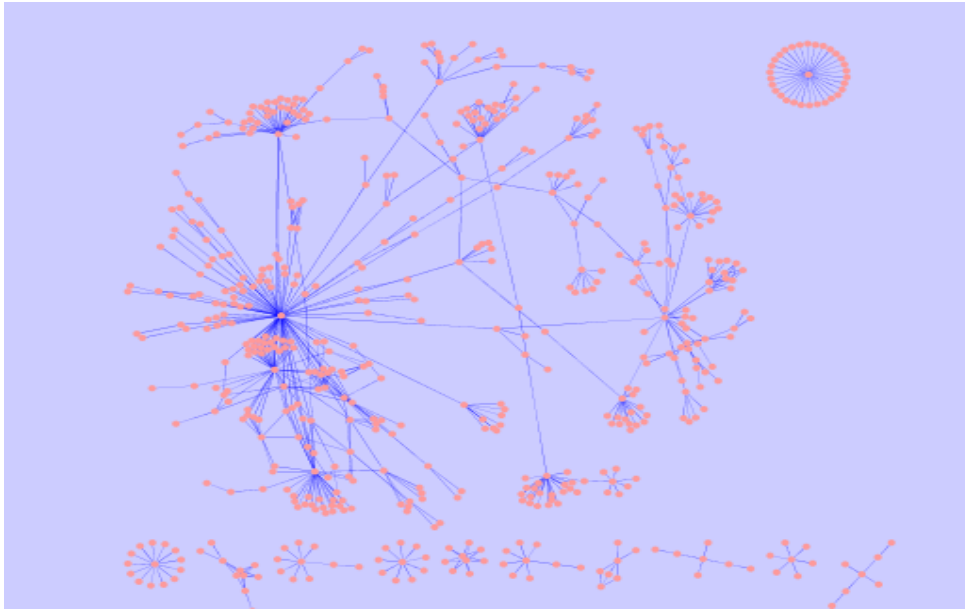
In questo capitolo si analizza una rete formata da due tipi di nodi: utenti e *hashtag*. Tutti gli *hashtag* sono collegati con gli utenti dai quali sono stati adoperati. In questa sezione saranno mostrate le reti risultanti divise per mesi e poi infine la totale di queste. I punti d'analisi saranno alcune misurazioni basilari delle reti (Es :grado, centralità) e alcune applicazioni di misure base a nodi ritenuti interessanti. Prendendo per esempio la distanza degli altri nodi dal nodo *KeyWord*.

Gli obiettivi sono:

- Mostrare quali sono gli *hashtag* e/o gli utenti più importanti in base alle misurazioni fatte;
- Comprendere quali sono i nodi più importanti per la struttura della rete;
- Ottenere maggiori informazioni su quali nodi siano vicini al nodo #autogrill.

⁸ -Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach; Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, Ming Zhang
-A new perspective on Twitter hashtag use: diffusion of innovation theory; Hsia Ching Ching Chang

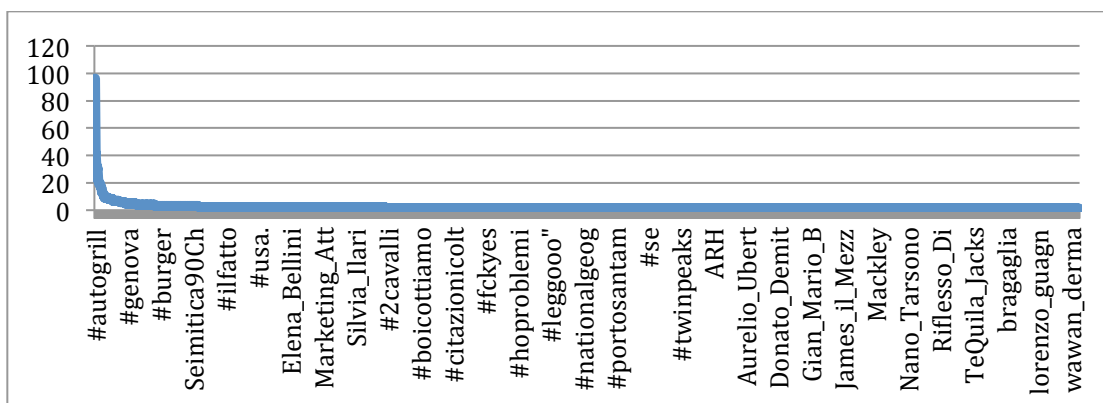
Aprile



I dati di base

| | |
|--------------------------------------|-------------|
| Numero di Nodi | 1308 |
| Numero di Archi | 1160 |
| Numero di Componenti | 298 |
| Diametro | 15 |
| Numero medio di vicini | 1,17 |
| Percorsi più brevi | 222528(13%) |
| Coefficiente di clustering | 0,00 |
| Lunghezza caratteristica dei cammini | 5,634 |

Distribuzione del grado



I primi 20

| | |
|--------------------|----|
| #autogrill | 96 |
| #benetton | 42 |
| #ultimaparola | 33 |
| #benetton) | 32 |
| #ibnux | 30 |
| #eccheccazzo | 22 |
| #camogli | 19 |
| #sapevatelo | 19 |
| Simone_Valtieri | 17 |
| Giulia_Brambilla | 16 |
| #devemorire | 13 |
| L'isola_dei | 12 |
| #spizzico | 10 |
| #9thapril2011 | 9 |
| #autogrill? | 9 |
| #donnaifaalcostumi | 9 |
| Addotta_Domenico | 9 |
| Dafne | 9 |
| #breakingitaly: | 8 |
| #burgerking | 8 |

Tra i primi venti non sorprende assolutamente #autogrill in prima posizione, in seconda troviamo #benetton , che la famiglia proprietaria di Autogrill. In terza posizione troviamo #ultimaparola che è riferito al programma Rai, questo può indicare che il programma tv abbia trattato argomenti molto vicini al marchio Autogrill. Eseguendo una ricerca si nota che il programma ha trattato in Aprile dei licenziamenti all'interno della società Autogrill; questo argomento nel mese di Aprile è stato molto sentito dalla comunità di Twitter, questo fatto è confermato anche dall'analisi dei *retweet* nel mese di Aprile(affrontato nel capitolo "la mia attrezzatura" nella sezione grafi).

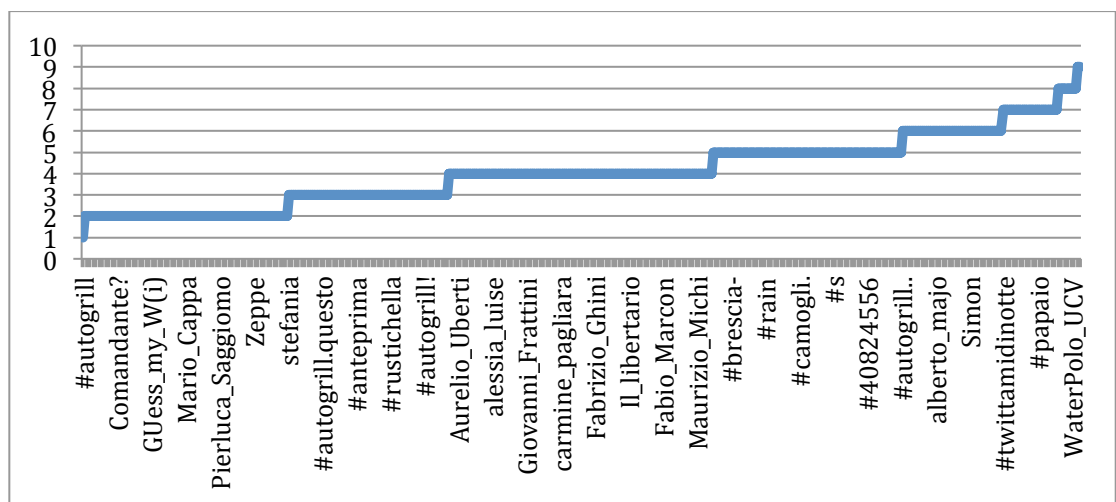
Centralità

| | |
|----------------------|------------------|
| #autogrill | 0.106994372931 |
| #camogli | 0.0443628576702 |
| Leonardo_Martinez | 0.0380897674915 |
| Expo_2015_Contact | 0.0265198340872 |
| #sapevatelo | 0.0260646719822 |
| Francesco_de_Frances | 0.0194241854999 |
| #benetton) | 0.0187067509575 |
| Daily_Tweetter_News | 0.0150315661709 |
| #eccheccazzo | 0.0149595431141 |
| #ultimaparola | 0.0145764319517 |
| #milano | 0.0137327897364 |
| Giulia_Brambilla | 0.0107456039369 |
| Tommaso_Innocenzi | 0.010093854308 |
| #benetton | 0.00945381457398 |
| #burgerking | 0.00918392366658 |
| Simone_Valtieri | 0.00897592265275 |

| | |
|---------------------|------------------|
| Alessia_Scurati | 0.00884073183297 |
| Il_Popolo_Viola | 0.00865935336006 |
| Pierluca_Saggiomo_ | 0.00865935336006 |
| Instagramers_Genova | 0.00751991555592 |

Per la centralità #autogrill la fa da padrone, ma anche #camogli ha un ruolo molto importante, si noterà nelle prossime misurazioni che #camogli tende generalmente in tutti i mesi ad essere un nodo molto centrale. Bisogna tener presente che Camogli è anche il nome di una città e della sua squadra di calcio il che lo rende un nodo importante anche al di fuori del mondo della ristorazione.

Distanza da Autogrill



| I 20 più lontani | | I 20 più vicini | |
|--------------------|---|---------------------|---|
| #google | 8 | #autogrill | 0 |
| #nu | 8 | ...:Anjia:... | 1 |
| BlazeSan | 7 | Addotta_Domenico | 1 |
| Danilo_Vallelonga | 7 | Ale | 1 |
| Francesca_Biscardi | 7 | Alessandro_Marra | 1 |
| Giuseppe_Macrv" | 7 | Alessandro_Proietti | 1 |
| Monica | 7 | Alessio_Marrari | 1 |

| | | | |
|--------------------|---|----------------------|---|
| TWITTAMIDINOTTE | 7 | Alex_Achille_ | 1 |
| WaterPolo_UCV | 7 | Amelya_Sachs | 1 |
| FrancescoMartucci | 7 | Andrea_Di_Turi | 1 |
| vtest2 | 7 | Andrea_Favaro | 1 |
| #twittamidinotte | 6 | Angelo_Mazzi | 1 |
| #instaday | 6 | Around_Food! | 1 |
| #instagramer | 6 | Bambini_Cardiopatici | 1 |
| #instagramhub | 6 | Benedetta_Miele | 1 |
| #instatalent | 6 | Carmelo | 1 |
| #iphoneography | 6 | Comandante?deCHE? | 1 |
| #iphonesi | 6 | Daniela_Maranzano | 1 |
| #garliccheesebread | 6 | Daniele | 1 |
| #italia | 6 | Daniele_Keshk | 1 |

PageRank dei nodi

I 20 migliori (in ordine crescente)

| | |
|----------------------|------------------|
| Francesco_de_Frances | 0.00249591754474 |
| #foreverpizza | 0.00253054279181 |
| #mistero | 0.00253054279181 |
| Benedetta_Joppolo | 0.00263350703752 |
| #twittamidinotte | 0.00266852169009 |
| Alessia_Scurati | 0.00267032813101 |
| #trotanewjob | 0.00281384006159 |
| Addotta_Domenico | 0.00309637457569 |
| #breakingitaly: | 0.00316971580441 |
| #9thapril2011 | 0.00358839264188 |
| #sapevatelo | 0.00494552698126 |
| #devemorire | 0.00499885910864 |

| | |
|------------------|------------------|
| Giulia_Brambilla | 0.00537863634637 |
| #camogli | 0.00551525385733 |
| Simone_Valtieri | 0.00583517220915 |
| #eccheccazzo | 0.00773902859106 |
| #ultimaparola | 0.00877723392885 |
| #benetton | 0.00880039405418 |
| #benetton) | 0.0104745967292 |
| #ibnux | 0.0109933415924 |
| #autogrill | 0.0237047888165 |

I 20 peggiori

| | |
|----------------------|-------------------|
| Jess | 0.000292512593616 |
| hesundermyskin_àû | 0.000292512593616 |
| Vanda_Focarini | 0.000293222171633 |
| #camogli!! | 0.000324727475304 |
| #dafne | 0.000324727475304 |
| #fiemme. | 0.000324727475304 |
| ...:Anjia:... | 0.000325005109532 |
| Alessandro_Marra | 0.000325005109532 |
| Alessandro_Proietti | 0.000325005109532 |
| Andrea_Favaro | 0.000325005109532 |
| Around_Food! | 0.000325005109532 |
| Bambini_Cardiopatici | 0.000325005109532 |
| Benedetta_Miele | 0.000325005109532 |
| Daniele_Scanferla | 0.000325005109532 |
| Davide_Marcucci | 0.000325005109532 |
| EleonoraSunshineGrr_ | 0.000325005109532 |
| GUess_my_W(i)osh | 0.000325005109532 |
| Gianluca_Droghetti | 0.000325005109532 |
| Giudici_Carlo | 0.000325005109532 |
| Mackley | 0.000325005109532 |

Per il mese di aprile ci troviamo davanti una rete di 1308 nodi e 1160 archi. Il numero di componenti connesse è di 298, un numero accettabile per le dimensioni e la tipologia della rete.

Il coefficiente di *clustering* è 0, questo tipo di rete (utente-*hashtag*) è una caratteristica costante.

La distribuzione del grado è una *Power-Law*⁹ standard, questo mostra che la rete è matura.

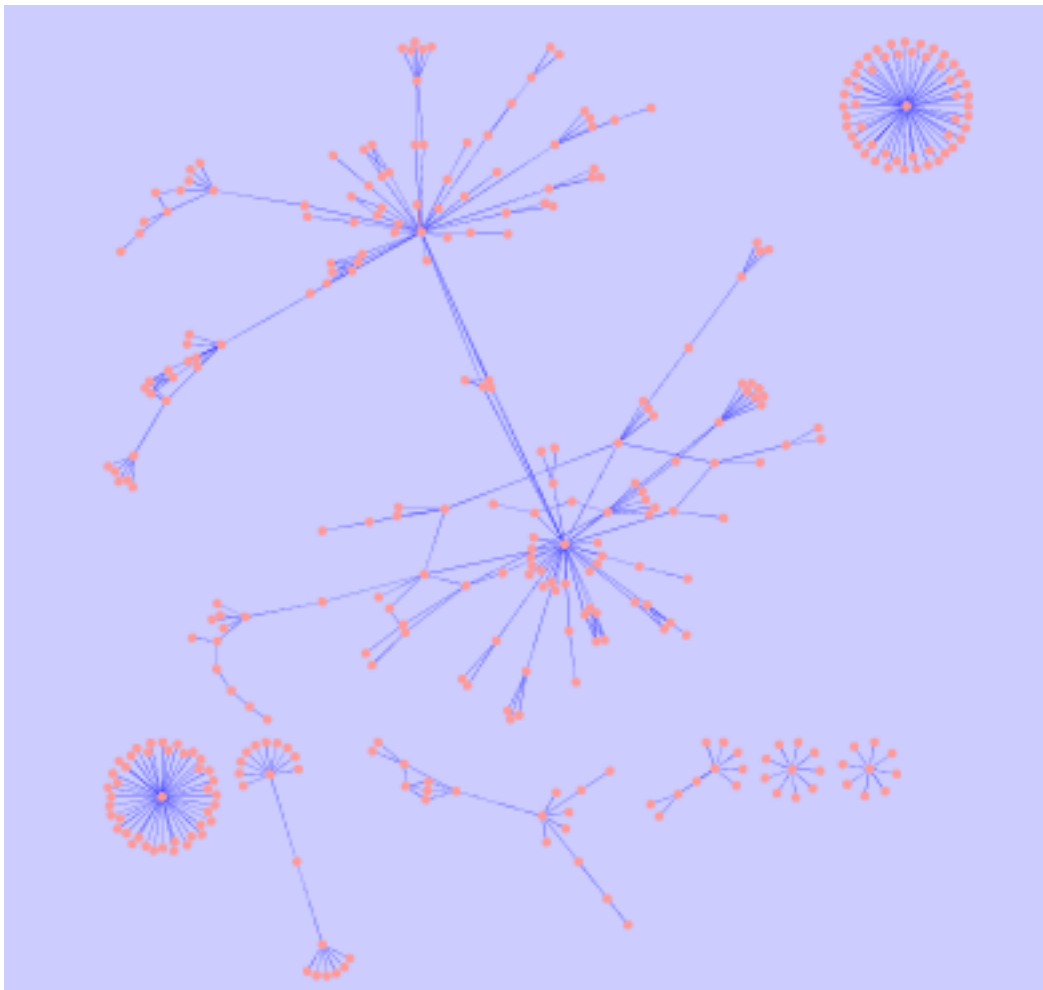
La classifica per centralità vede la supremazia di #autogrill su un secondo #camogli, e #benetton scivola sesto.

Tra i nodi più lontani da #autogrill troviamo #google. Tra quelli più vicini invece troviamo che i primi 20 risultati sono solo utenti, come è normale in questa tipologia di rete. A distanza 1 da un nodo *hashtag* ci saranno gli utenti che hanno utilizzato tale hashtag e l'utente è il mezzo di collegamento tra gli hashtag, quindi è improbabile trovare due *hashtag* con distanza 1 tra di loro.

In base al PageRank gli *hashtag* più importanti sono: #autogrill, #ibnux, #benetton, #ultimaparola. #ibnux è un riferimento a Orang Kampung, un utente Twitter molto attivo.

⁹ http://it.wikipedia.org/wiki/Legge_di_potenza

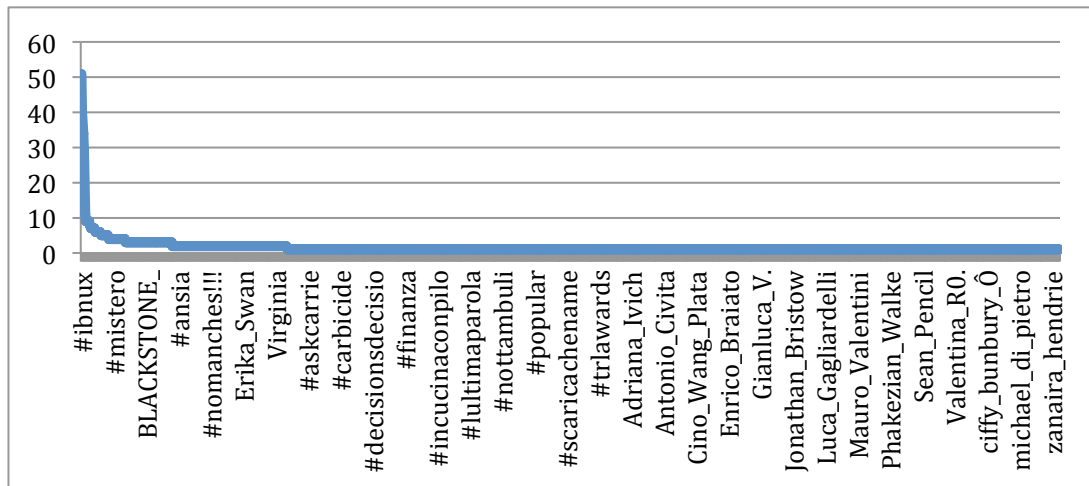
Maggio



I dati di base

| | |
|--------------------------------------|-----------|
| Numero di Nodi | 1035 |
| Numero di Archi | 797 |
| Numero di Componenti | 288 |
| Diametro | 15 |
| Numero medio di vicini | 1,15 |
| Percorsi più brevi | 41986(3%) |
| Coefficiente di clustering | 0,00 |
| Lunghezza caratteristica dei cammini | 4,623 |

Distribuzione del grado



I primi 20

| | |
|----------------------|----|
| #ibnux | 51 |
| #ff?" | 40 |
| #camogli | 34 |
| #autogrill | 29 |
| #mannarinoblog | 11 |
| #tiauguro | 9 |
| ANDREA_ALVIGINI | 9 |
| Beatrice_Lombardo | 9 |
| Davide_Ghidini | 9 |
| #pizza | 8 |
| #allucinazioni | 7 |
| #filmgeografici | 7 |
| #sapevatelo | 7 |
| Locanda_I_Tre_Merli_ | 7 |
| Naomi_Ludmilla_R. | 7 |
| #cv®lacrisidelcinema | 6 |
| #lol | 6 |
| #salto12 | 6 |
| I_Blog_del_Fatto | 6 |
| Vincenzo_ | 6 |

#Camogli e #Autogrill appaiono in meno messaggi di quanto non fosse nel mese precedente, e vediamo comparire nelle prime posizioni #ff che nel gergo di Twitter significa *Follow Friday*¹⁰ ed è una raccomandazione a seguire una determinata persona o argomento. Il suo accostamento dovuto al termine Autogrill è puramente ironico, infatti il Tweet originale è:

« *“Cosa è un #ff?” E’ quando qualcuno consiglia gente che merita, ai miei tempi scrivevamo il i numeri di telefono nei bagno dell’Autogrill»*

Centralità

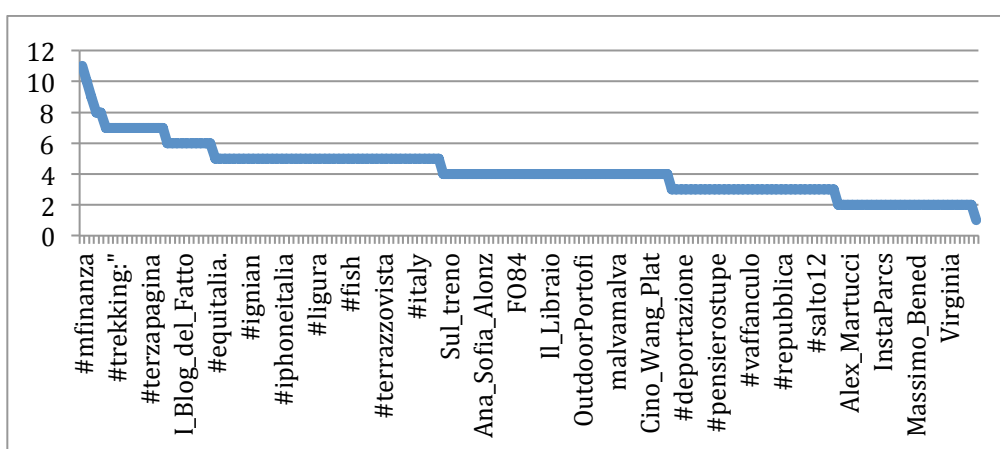
I top 20

| | |
|----------------------|------------------|
| #camogli | 0.0240303074828 |
| #autogrill | 0.0227340333947 |
| ,ôî_Julianne_Marie | 0.00600345222925 |
| ANDREA_ALVIGINI | 0.00563568951761 |
| Dissapore_Media | 0.00548238565391 |
| Luca_Brazzoli | 0.00548238565391 |
| gillian__ | 0.00548238565391 |
| Locanda_I_Tre_Merli_ | 0.00432478505047 |
| #blog | 0.0036811834239 |
| I_Blog_del_Fatto | 0.00342215819068 |
| Roberto_Reano | 0.00336543733669 |
| #salto12 | 0.00308939584728 |
| Davide_Ghidini | 0.00258725873155 |
| Beatrice_Lombardo | 0.00255376641776 |
| #ibnux | 0.00241063629454 |
| #autogrill. | 0.00240118281888 |
| #sapevatelo | 0.00174983834557 |

¹⁰ <http://www.mediaroyal.com/2010/02/followfriday-su-twitter/>

| | |
|------------|------------------|
| Vincenzo_ | 0.001748892998 |
| Federica | 0.00174700230287 |
| #mannarino | 0.0017375488272 |

Distanza da Autogrill



| I 20 più lontani | | I 20 più vicini | |
|---------------------|----|--------------------|---|
| #mfinanza | 10 | #autogrill | 0 |
| Quotidiani_Italiani | 9 | InstaParcs | 1 |
| #ilfatto | 8 | Jane_Floris | 1 |
| Alessandra | 7 | Largo_Consumo | 1 |
| Chiara_Cancellieri | 7 | Linkontro_Nielsen | 1 |
| #camogli! | 6 | Luca_Brazzoli | 1 |
| #trail | 6 | Marco_Nurra | 1 |
| #trekking:" | 6 | Massimiliano_Nessi | 1 |
| #hitotit | 6 | Massimo_Benedetti | 1 |

| | | | |
|----------------|---|-------------------|---|
| #vara | 6 | Matteo_Radice | 1 |
| #allucinazione | 6 | Mauro_Strapparava | 1 |
| #crisi | 6 | Michele_Azzu | 1 |
| #mannarino | 6 | Peppe_Marici | 1 |
| #paesereale | 6 | Roberto_Parodi | 1 |
| #terzapagina | 6 | Roberto_Reano | 1 |
| #cronaca | 6 | Virginia | 1 |
| #china | 6 | gillian__ | 1 |
| #photo | 6 | lagonzi | 1 |
| Arbaspv†a_Tour | 5 | raven | 1 |
| Liguria_HitHot | 5 | roberta_carbone | 1 |

Da questa classifica si nota una cosa singolare, i nodi riguardanti argomenti “autorevoli” come #mfinanza (Milano finanza), #ilfatto e Quotidiani italiani rimangano decisamente lontani da Autogrill. Questo vuol dire che l’azienda su Twitter non è solo finanza, anzi si guarda poco a quest’ aspetto.

PageRank dei nodi

Top 20 (ordine ascendente)

| | |
|----------------------|------------------|
| nawin_sangpi | 0.00230910524884 |
| #salto12 | 0.00250297553063 |
| Vincenzo_ | 0.00261331271717 |
| I_Blog_del_Fatto | 0.00269333387363 |
| Locanda_I_Tre_Merli_ | 0.00274945170159 |
| #sapevatelo | 0.0029944571633 |
| #filmgeografici | 0.00304830095434 |
| ANDREA_ALVIGINI | 0.00311876803401 |
| #cv®lacrisidelcinema | 0.00320125955713 |
| #lol | 0.00320125955713 |

| | |
|-------------------|------------------|
| #allucinazioni | 0.00340865262557 |
| Beatrice_Lombardo | 0.00347830814661 |
| Naomi_Ludmilla_R. | 0.00350985954393 |
| Davide_Ghidini | 0.00381136349094 |
| #pizza | 0.00409341386542 |
| #tiauguro | 0.00453949101956 |
| #mannarinoblog | 0.00522425225941 |
| #autogrill | 0.0101021592585 |
| #camogli | 0.0121184201342 |
| #ff?" | 0.018367882798 |
| #ibnux | 0.0232747314936 |

I 20 peggiori

| | |
|--------------------|-------------------|
| #linkontronielsen: | 0.000429116110895 |
| #caffebenecomune | 0.000440181381581 |
| #ultimaparola | 0.000440181381581 |
| #mavaffanculova | 0.000440181381581 |
| #sciacalli | 0.000440181381581 |
| #ultimaparola | 0.000440181381581 |
| Alex_Martucci | 0.000441728839651 |
| Giovanna_Leoni | 0.000441728839651 |
| Massimo_Benedetti | 0.000441728839651 |
| Matteo_Radice | 0.000441728839651 |
| Mauro_Strapparava | 0.000441728839651 |
| Peppe_Marici | 0.000441728839651 |
| Alessio_Uda | 0.000448591572179 |
| AndreaCamperi | 0.000448591572179 |
| Cindy_Giuffrida | 0.000448591572179 |
| Edoardo_Portaluppi | 0.000448591572179 |
| FO84 | 0.000448591572179 |

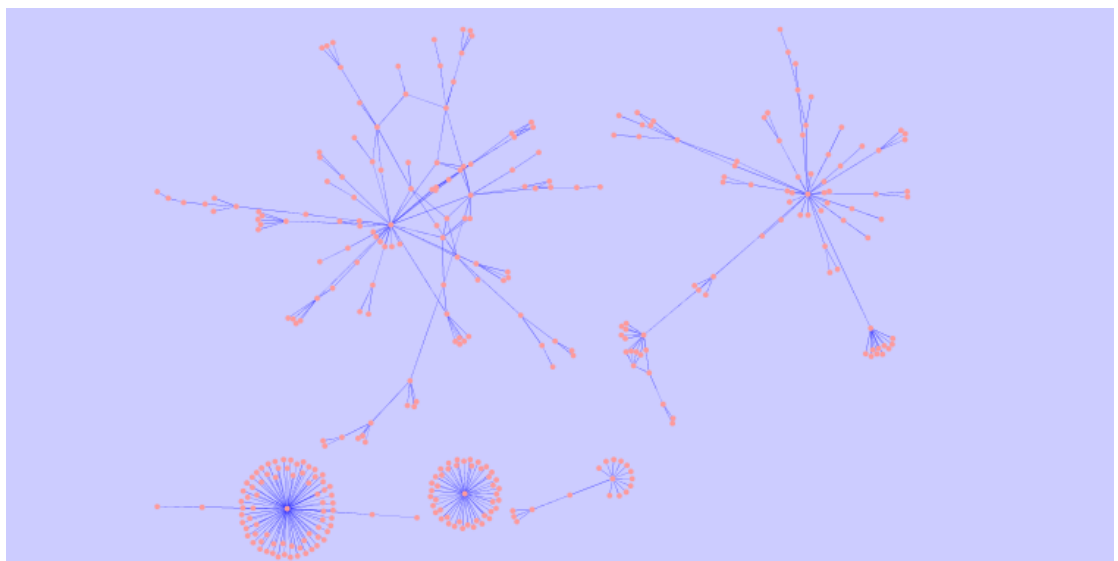
| | |
|---------------------|-------------------|
| Francesco_Impallari | 0.000448591572179 |
| Giorgia_Borneto | 0.000448591572179 |
| Hotel_Crow | 0.000448591572179 |

La rete del mese di maggio è composta da 1035 nodi e 797 archi. Il numero di componenti è di 288. La distribuzione del grado è simile al mese precedente, per lo meno nell'aspetto grafico. I nodi con grado più alto sono: #ibnux, #ff?, # camogli, #autogrill.

In base alla centralità i nodi più importanti sono #camogli e #autogrill; come era già stato detto #camogli tende ad essere un nodo molto centrale e questo non era prevedibile a priori, invece la centralità alta di #autogrill è quasi scontata visto che è al centro della ricerca.

I nodi più distanti da #autogrill sono #mfinanza(10), quotidiani italiani (9), #ilfatto(8). In base al *PageRank* i più importanti sono #ibnux, #ff?, #camogli e #autogrill, rispecchiando quindi i risultati del grado.

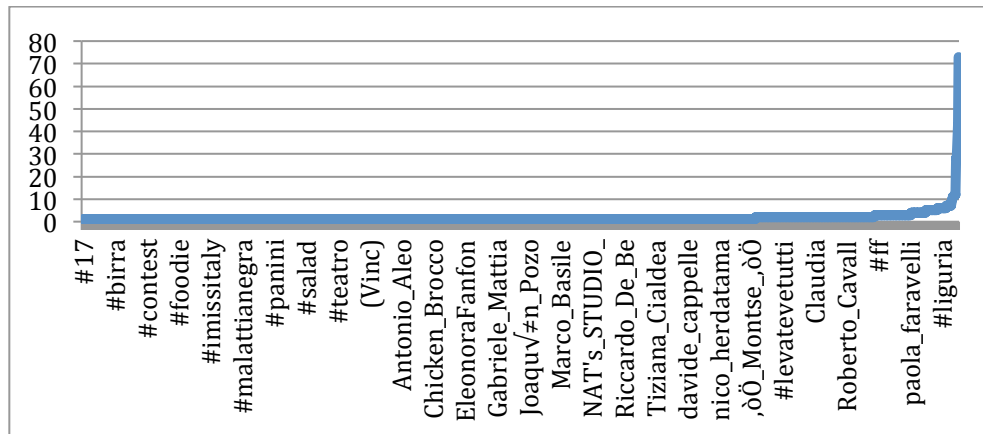
Giugno



I dati di base

| | |
|--------------------------------------|-----------|
| Numero di Nodi | 1081 |
| Numero di Archi | 869 |
| Numero di Componenti | 264 |
| Diametro | 12 |
| Numero medio di vicini | 1,6 |
| Percorsi più brevi | 31292(2%) |
| Coefficiente di clustering | 0,00 |
| Lunghezza caratteristica dei cammini | 3,53 |

La distribuzione del grado



I primi 20

| | |
|------------------|----|
| #eurofighi | 73 |
| #ibnux | 43 |
| #autogrill | 28 |
| #camogli | 28 |
| #ballarò | 12 |
| #iaskq" | 12 |
| Addotta_Domenico | 11 |
| risoltoilmistero | 11 |
| #italiacroazia | 10 |
| #50cosasqueamo | 7 |
| #buongiorno | 7 |

| | |
|-----------------------------|---|
| #euro2012 | 7 |
| #isolitiidioti | 7 |
| #onu | 7 |
| #siria | 7 |
| Leonardo_Russo_Spena | 7 |
| #camogli, | 6 |
| #concertoemilia | 6 |
| #italyabsolutelyneedstaylor | 6 |
| #lamorte | 6 |
| #liguria | 6 |

Centralità

I Top 20

| | |
|----------------------|-------------------|
| #camogli | 0.00935566732864 |
| #autogrill | 0.00575548915162 |
| #eurofighi | 0.00480360313542 |
| risoltoilmistero | 0.0033748955994 |
| Roberto_Miliacca | 0.00229353427742 |
| #instagram | 0.00213012284729 |
| #ballarò | 0.00198345675804 |
| Desak_Yuliantari | 0.00189511064917 |
| #camogli, | 0.00159454827793 |
| #ibnux | 0.00156424581006 |
| Leonardo_Russo_Spena | 0.00146273439868 |
| apolide_sardo | 0.00145454320484 |
| Addotta_Domenico | 0.00139448269889 |
| Aurora_Incardona | 0.00133385301633 |
| #pizza | 0.00117448356503 |
| #liguria | 0.0011733452118 |
| #madonna | 0.00116928673509 |
| #sapevatelo | 0.000987397687411 |

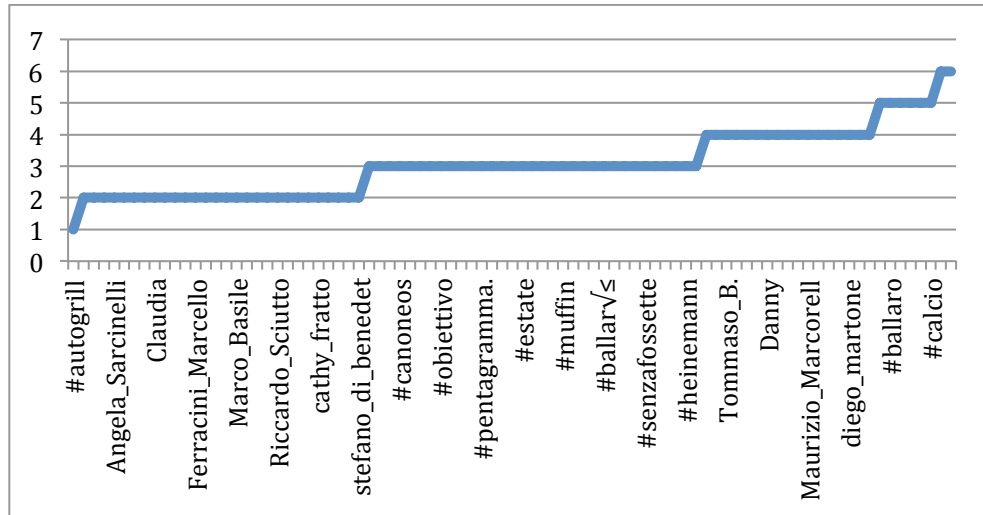
Valeria_Abate

0.000987397687411

#genova

0.000980468580832

Distanza da Autogrill



| I 20 più lontani | | I 20 più vicini | |
|----------------------|---|----------------------|---|
| Claudio_Lauretta | 5 | #autogrill | 0 |
| Michela_Daglio | 5 | Addotta_Domenico | 1 |
| #maancheno | 4 | Alberta_Tresoldi | 1 |
| #ballaro | 4 | Alo_B. | 1 |
| #tergicristallo | 4 | Angela_Sarcinelli | 1 |
| #occupyautogrill | 4 | Antonietta_Uliano | 1 |
| #ff | 4 | Biljana_Pucar | 1 |
| #calcio | 4 | Blitz_quotidiano | 1 |
| Movimento_Anti_Casta | 3 | Claudia | 1 |
| RoccoZaroOriginal | 3 | Coca-Cola_HBC_Italia | 1 |
| Tommaso_B. | 3 | Esprit | 1 |
| abidibo | 3 | Fabrizio_Benzoni | 1 |
| AndreaSpinosiPicotti | 3 | Ferracini_Marcello | 1 |
| Angela75 | 3 | Giacomo_Paci | 1 |
| Danny | 3 | loska_Versari | 1 |

| | | | |
|---------------------|---|---------------------|---|
| EleonoraFanfoni | 3 | Luca_Pellegrini | 1 |
| Luca_Bersaglia | 3 | Marco_Basile | 1 |
| Matteo_De_Felice | 3 | Monica_Berardinelli | 1 |
| Maurizio_Marcorelli | 3 | Monica_Dalla_Corte | 1 |
| Miguel_Vargas_A. | 3 | Raibaz | 1 |

PageRank sui nodi

I 20 più importanti(ordine crescente)

| | |
|------------------------------|------------------|
| #masdar | 0.00263739578508 |
| #facebook | 0.00263739578508 |
| #iaskq,Äù | 0.00263739578508 |
| Leonardo_Russo_Spena | 0.00265902617391 |
| Maranatha_ITALY | 0.00288520281642 |
| #euro2012 | 0.00300242219272 |
| #concertoemilia | 0.00306440272391 |
| #tweetyourfavoritepizzaplace | 0.00306440272391 |
| #italyabsolutelyneedstaylor | 0.00306440272391 |
| risoltoilmistero | 0.00326976499785 |
| #buongiorno | 0.00335101869059 |
| #50cosasqueamo | 0.00349140966274 |
| #isolitiidioti | 0.00349140966274 |
| #ballarò | 0.00431049212323 |
| #italiacroazia | 0.00477243047925 |
| Addotta_Domenico | 0.00482618550224 |
| #iaskq" | 0.00545664761334 |
| #camogli | 0.00907353855992 |
| #autogrill | 0.0109334951598 |

| | |
|------------|-----------------|
| #ibnux | 0.0188636594608 |
| #eurofighi | 0.0315674568563 |

I 20 meno importanti

| | |
|----------------------|-------------------|
| #igersliguria | 0.000413653350725 |
| Adriano_Nuzzo | 0.000414851912058 |
| Anna_Ligabue | 0.000414851912058 |
| Federico_Sandi | 0.000414851912058 |
| Laura_Carignani | 0.000414851912058 |
| M_qualunque | 0.000414851912058 |
| Marco_Scotti | 0.000414851912058 |
| #fotografa | 0.000423997907366 |
| Chiaradevito | 0.000433136778669 |
| Carlotta_Grazioli | 0.000434343882423 |
| AndreaSpinosiPicotti | 0.000444731728148 |
| Danny | 0.000444731728148 |
| EleonoraFanfoni | 0.000444731728148 |
| Matteo_De_Felice | 0.000444731728148 |
| ardovig | 0.000444731728148 |
| diego_martone | 0.000444731728148 |
| #camogli! | 0.000448084115862 |
| #viaggio | 0.000455358835862 |
| #boat | 0.000462286952832 |
| #instacool | 0.000462286952832 |

Nel mese di giugno ci sono 1081 nodi e 869 archi. Il numero di componenti connesse si attesta a 264.

I quattro nodi con il grado più alto sono #eurofighi, #ibnux, #autogrill e #camogli, anche se nelle posizioni più basse si trova #ballarò il quale è collegato al famoso programma Rai.

«#ballarò ha fatto vedere gli Autogrill deserti alle 13:0»>

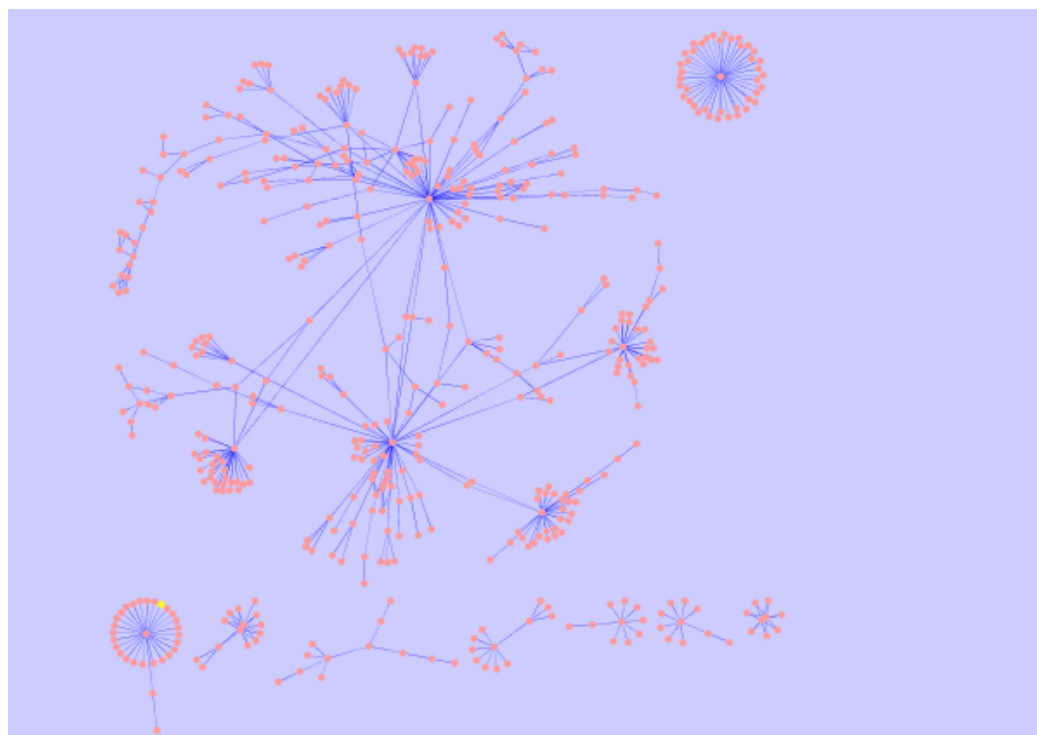
Il servizio di Ballarò utilizza l'immagine dell'autogrill deserto per parlare della crisi¹¹.

In base alla centralità la classifica è: #camogli, #autogrill, #eurofighi. Gli *hashtag* più lontani da #autogrill sono: #maancheno, #ballaro, #tergicristallo, #occupyautogrill.

In base al *PageRank* gli *hashtag* più importanti sono #eurofighi, #ibnux, #autogrill e #camogli . Appare anche #masdar che è una città in costruzione vicino ad Abu Dhabi e viene associata ad Autogrill con il seguente tweet:

« *In attesa di #Masdar City fate, un salto all'Autogrill Villoresi.....* »

Luglio

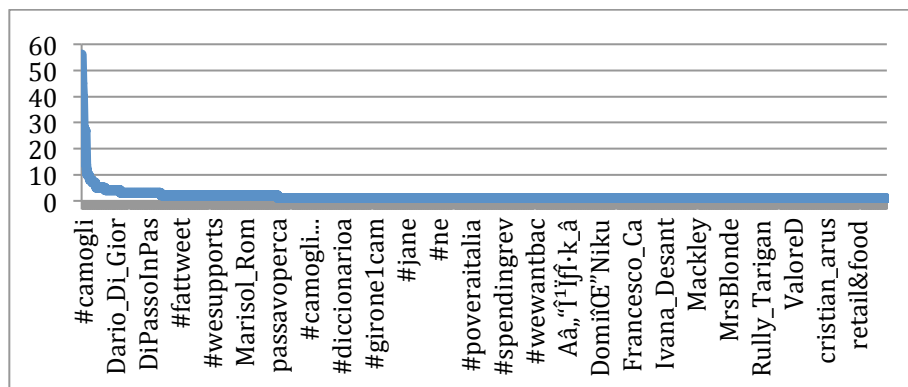


¹¹ <http://www.rai.tv/dl/RaiTV/programmi/media/ContentItem-1b06ff48-0cb7-43a4-9700-d873332ba323.html>

I dati di base

| | |
|--------------------------------------|------------|
| Numero di Nodi | 1253 |
| Numero di Archi | 1020 |
| Numero di Componenti | 308 |
| Diametro | 17 |
| Numero medio di vicini | 1,62 |
| Percorsi più brevi | 128910(8%) |
| Coefficiente di <i>clustering</i> | 0,00 |
| Lunghezza caratteristica dei cammini | 5,574 |

Distribuzione del grado



I primi 20

| | |
|---------------|----|
| #camogli | 56 |
| #autogrill | 45 |
| #ibnux | 40 |
| #saratommasi | 28 |
| Valeria_Abate | 28 |
| #anagrammi | 27 |
| #sapevatelo | 27 |

| | |
|-------------------|----|
| #liguria | 14 |
| valentina_rimondi | 12 |
| #gossip | 10 |
| TARI | 10 |
| #youp | 9 |
| #querelaconpaola | 8 |
| #tommasi | 8 |
| #wma | 8 |
| andrea_vetralla | 8 |
| #ff | 7 |
| #wild | 7 |
| Adriano_Gimelli | 7 |
| Libri_Libri_Libri | 7 |

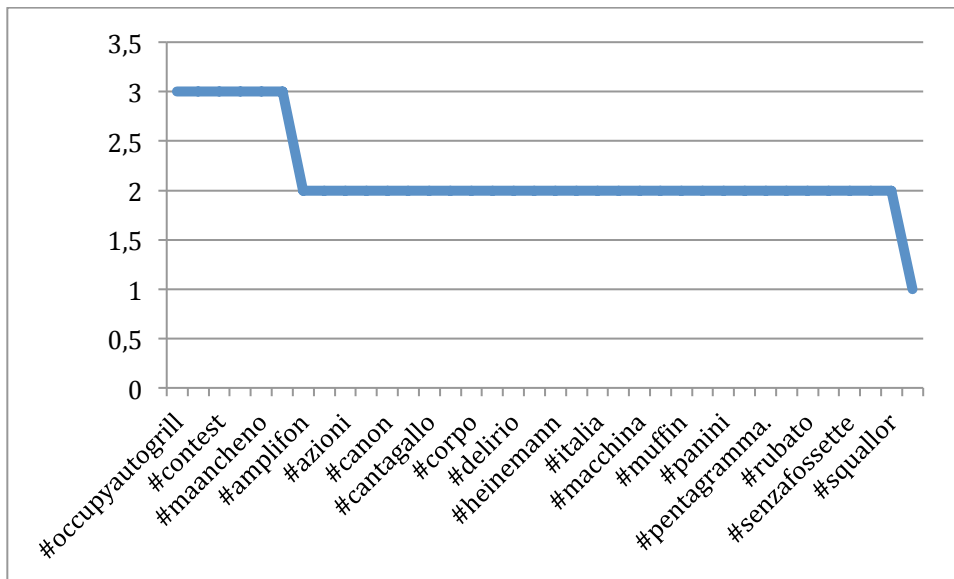
In questo mese #camogli ed #autogrill tornano in cima alle classifica per il grado; in questo periodo ci sono stati molti *tweet* comici che collegano Autogrill a Sara Tommasi. L'ironia dimostra spesso la capacità di potersi diffondere velocemente su Twitter.

Centralità

| | |
|--------------|------------------|
| #camogli | 0.0489588300747 |
| #liguria | 0.0205054840248 |
| #autogrill | 0.0198818825428 |
| #instagram | 0.0119401342577 |
| #camogli, | 0.0101610359121 |
| #genova | 0.00623601481971 |
| #madonna | 0.00418179817321 |
| #pizza | 0.00418179817321 |
| #portofino. | 0.00418179817321 |
| #igersgenova | 0.00393113483242 |

| | |
|-----------------|-------------------|
| #sapevatelo | 0.00363156157148 |
| #autogrill, | 0.00212758152672 |
| #ballarò | 0.00124720296394 |
| #italia | 0.00124720296394 |
| #promo | 0.00124720296394 |
| #coloreitaliano | 0.000794786202512 |
| #spizzico | 0.000293459520927 |
| #photography | 0.00018341220058 |
| #euro2012 | 0.000110047320348 |
| #autogrill... | 7,34E+06 |

Distanza da Autogrill



| I 20 più vicini | | I 20 più lontani | |
|-----------------|---|------------------|----|
| #autogrill | 0 | #arte | 12 |
| #amplifon | 1 | #damienhirst | 12 |
| #articolo | 1 | #app | 12 |

| | | | |
|------------------|---|--------------------|----|
| #azioni | 1 | #bkmobilecrowncard | 12 |
| #ballarò | 1 | #fastfood | 12 |
| #canon | 1 | #payment: | 12 |
| #canoneos | 1 | Adriano_D | 11 |
| #cantagallo | 1 | Mister_Forky | 11 |
| #cinesi; | 1 | The_Lazy_Guy | 11 |
| #corpo | 1 | ViviLondra.it | 11 |
| #crisi: | 1 | #burgerking | 10 |
| #delirio | 1 | Pierangelo_Ranieri | 9 |
| #estate | 1 | #fail | 8 |
| #heinemann | 1 | #fired | 8 |
| #inedrustichella | 1 | #firenze | 8 |
| #italia | 1 | #ilg | 8 |

PageRank

I top 20 (ordine decrescente)

| | |
|---------------|------------------|
| #articolo | 0.00592362595037 |
| #canon | 0.00592362595037 |
| #mia | 0.00592362595037 |
| #rubato | 0.00592362595037 |
| #canoneos | 0.00592362595037 |
| #macchina | 0.00592362595037 |
| #obiettivo | 0.00592362595037 |
| #corpo | 0.00592362595037 |
| #abbazia | 0.00621046652335 |
| #autogrill... | 0.00624231263209 |
| #pizzahut | 0.00624231263209 |
| #sapevatelo | 0.00686622368187 |

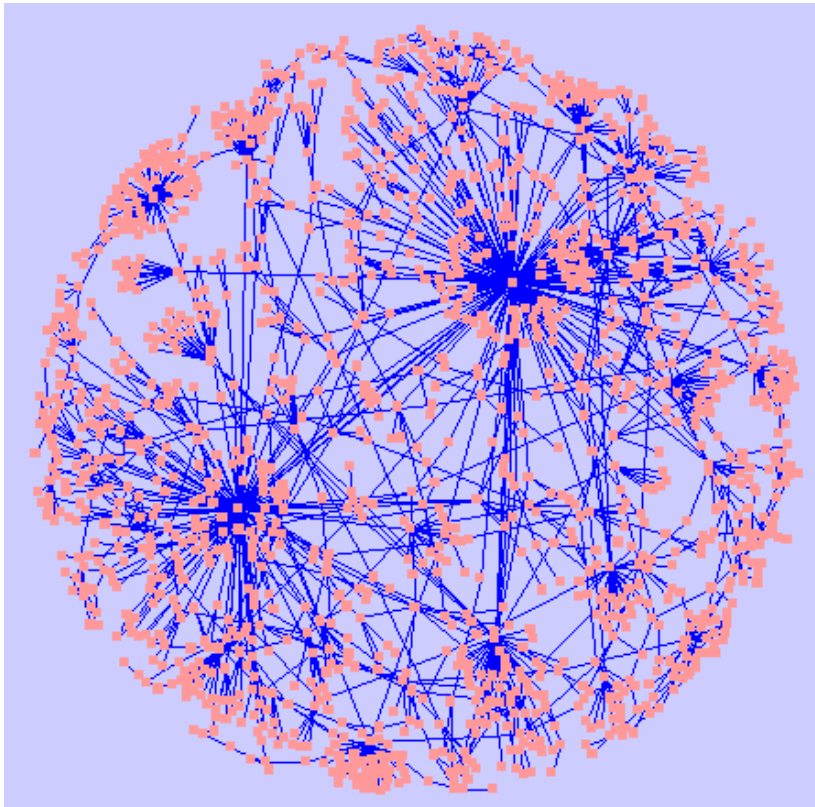
| | |
|--------------|------------------|
| #genova | 0.00730551645573 |
| #igersgenova | 0.00777984348567 |
| #euro2012 | 0.00816561389777 |
| #spizzico | 0.00869387418065 |
| #camogli, | 0.00888316585898 |
| #liguria | 0.010141689709 |
| #instagram | 0.0103212671989 |
| #autogrill | 0.0300109494657 |
| #camogli | 0.0322990543745 |

I 20 peggiori

| | |
|------------------|------------------|
| #habawaba | 0.00142270347857 |
| #liguria: | 0.00142270347857 |
| #teatrosociale | 0.00142270347857 |
| #waugh | 0.00142270347857 |
| #photo | 0.0014479446716 |
| #camogli! | 0.00150034149762 |
| #cantagallo | 0.00151792914978 |
| #delirio | 0.00151792914978 |
| #heinemann | 0.00151792914978 |
| #ineedrutchella | 0.00151792914978 |
| #panino?! | 0.00151792914978 |
| #pentagramma. | 0.00151792914978 |
| #politica | 0.0015253962991 |
| #occupyautogrill | 0.00159235769822 |
| #dukan.. | 0.00171696801242 |
| #mordiefuggi | 0.00171696801242 |
| #contest | 0.00183843454315 |
| #calcio | 0.00183843454315 |
| #ballaro | 0.00209737040473 |
| #maancheno | 0.00209737040473 |

Se si guardano i nodi più vicini ad Autogrill ci si accorge che questo mese il nodo #ballarò è a distanza 1, quindi molto vicino. Questo è dovuto al fatto che i due termini compaiono nel medesimo spesso nel medesimo *tweet*. Contrariamente però nella classifica del *PageRank* #ballarò rientra tra i peggiori, questo può essere dovuto al fatto che nonostante sia molto vicino ad Autogrill l'*hashtag* è collegato solo con quest'ultimo, infatti #ballarò non rientra nella top 20 in base al grado.

Totale



I dati di base

| | |
|--------------------------------------|--------------|
| Numero di Nodi | 3763 |
| Numero di Archi | 3297 |
| Numero di Componenti | 824 |
| Diametro | 20 |
| Numero medio di vicini | 1,752 |
| Percorsi più brevi | 2058400(14%) |
| Coefficiente di clustering | 0,00 |
| Lunghezza caratteristica dei cammini | 6,175 |

Grado

I top 20

| | |
|-----------------|-----|
| #autogrill | 190 |
| #camogli | 125 |
| #ibnux | 89 |
| #eurofighi | 67 |
| #sapevatelo | 53 |
| #benetton | 42 |
| #ff?" | 39 |
| #ultimaparola | 33 |
| #benetton) | 32 |
| Valeria_Abate | 31 |
| #saratommasi | 26 |
| #anagrammi | 25 |
| #liguria | 25 |
| #eccheccazzo | 21 |
| #ff | 17 |
| #spizzico | 17 |
| Simone_Valtieri | 17 |

| | |
|------------------|----|
| #autogrill. | 16 |
| #pizza | 16 |
| Giulia_Brambilla | 16 |

Centralità

| | |
|-------------------|------------------|
| #autogrill | 0.0990553940787 |
| #camogli | 0.0785312065047 |
| #ff?" | 0.0166038865997 |
| valentina | 0.0158469067263 |
| #sapevatelo | 0.0151766534313 |
| #eurofighi | 0.0147544863581 |
| Blitz_quotidiano | 0.0134194918632 |
| #benetton) | 0.0122206329795 |
| Valeria_Abate | 0.00986675342088 |
| Francesca | 0.00840942609659 |
| #instagram | 0.00828807010846 |
| #usa | 0.00823393045856 |
| Desak_Yuliantari | 0.00774170446315 |
| #pizza | 0.00717813247182 |
| Marco_Del_Sal | 0.00695284844801 |
| matteo_bandera | 0.00641565782299 |
| Leonardo_Martinez | 0.00634555591229 |
| Massimo_Scalzo | 0.00594268162142 |
| #ff | 0.00593623641261 |
| Il_Popolo_Viola | 0.00581519454214 |

Distanza da Autogrill

| I 20 più vicini | | I 20 più lontani | |
|---------------------|---|-------------------|----|
| #autogrill | 0 | #sanroccodi... | 10 |
| ...:Anjia:... | 1 | #serviceexport | 10 |
| ANDREA_ALVIGINI | 1 | #landscapes" | 10 |
| Addotta_Domenico | 1 | #barbeque | 10 |
| Alberta_Tresoldi | 1 | #dinner | 10 |
| Alberto_Depari_ | 1 | #smmasinag | 10 |
| Ale | 1 | #chocolatevolcano | 10 |
| Alessandro_Marra | 1 | #batam | 10 |
| Alessandro_Proietti | 1 | #beef | 10 |
| Alessio_Marrari | 1 | #beeffettucine | 10 |
| Alessio_Polastri | 1 | #fettucine | 10 |
| Alex_Achille_ | 1 | #instabatam | 10 |
| Alex_Martucci | 1 | #pasta | 10 |
| Alo_B. | 1 | #finedelmondo | 10 |
| Amelya_Sachs | 1 | #mayapensatecivoi | 10 |
| Andrea_Di_Turi | 1 | Monica_Tedde | 9 |
| Andrea_Favaro | 1 | offertelavoro | 9 |
| Andrea_Gori | 1 | Chloe | 9 |
| Angela_Sarcinelli | 1 | Il_Furgone_Di | 9 |
| Angelo_HIM_Russo | 1 | Emiliano | 9 |

Pagerank

I top 20 (ordine ascendente)

| | |
|------------------|------------------|
| #spizzico | 0.00149810976449 |
| Addotta_Domenico | 0.0015766387717 |
| #mistero | 0.00161226475538 |
| Giulia_Brambilla | 0.0016800568614 |

| | |
|-----------------|------------------|
| #ff | 0.00170622139075 |
| Simone_Valtieri | 0.00182594994766 |
| #pizza | 0.00183984756446 |
| #liguria | 0.00210730383722 |
| #eccheccazzo | 0.00246706903043 |
| #saratommasi | 0.0029583256799 |
| #ultimaparola | 0.00304057185637 |
| #benetton | 0.00315555427886 |
| #anagrammi | 0.00318570296484 |
| Valeria_Abate | 0.00349709916186 |
| #benetton) | 0.00354036833036 |
| #ff?" | 0.00463150030173 |
| #sapevatelo | 0.00532344195107 |
| #eurofighi | 0.0080831976888 |
| #ibnux | 0.0110339369512 |
| #camogli | 0.0110371860597 |
| #autogrill | 0.0173768480496 |

L'analisi del periodo totale ripropone ovviamente alcune cose che sono già state viste nelle reti dei singoli mesi, però in alcuni casi riesce a far capire il peso che alcuni eventi hanno avuto rispetto agli altri. Un particolare *hashtag* è #sapevatelo, che è nella maggior parte dei casi associato ad una critica. Questo *hashtag* si è già fatto notare nella rete di Aprile, ed è fortemente collegato con la protesta per i licenziamenti. Il fatto che il suo peso sia rilevante anche nella rete totale indica che le critiche non si sono arrestate, ma solo generalizzate. #autogrill ha una posizione importante, ed è ovvio il perché, #camogli invece si è dimostrato un perno della rete inaspettato. La presenza o l'assenza del nodo #camogli ha una rilevanza importantissima per la struttura della rete. Un altro nodo assai rilevante è #ibnux il quale meriterebbe un'analisi più approfondita ma che esula da questo lavoro. Tra le parole chiavi più utilizzate, si trova #benetton, il che sta a indicare che l'associazione tra il marchio Autogrill e la famiglia

Benetton è molto forte sulla rete. Nella classifica dei migliori venti in base al grado si nota #spizzico, il quale però sembra non avere un grande valore di centralità e per *Pagerank* si classifica 20°.

9. Classificare gli *Hashtag*: un approccio al *sentiment analysis* basato sul grafo

Nel precedente capitolo è stata analizzata la relazione tra utente e *hashtag*, in questo invece si riassume un'analisi portata avanti da alcuni studiosi sulle relazioni tra gli stessi *hashtag*.

L'obiettivo principale del lavoro di Wang¹² è quello di sviluppare e testare nuovi strumenti per il *sentiment analysis* su Twitter. L'analisi delle opinioni, su di un determinato *topic* (argomento), è uno degli obiettivi, ma soprattutto è una delle speranze, di chi deve comunicare sul web o di chi voglia monitorare la *percezione comune* di un determinato argomento.

Poniamo il caso della *Apple* che vuole sapere qual è il giudizio sul proprio prodotto, da parte del popolo di Twitter. Pensiamo al presidente Obama che vuole iniziare la sua futura campagna elettorale basandosi sulla rete. Sarebbe utilissimo in questi casi, per i diretti interessati, comprendere la percezione del "prodotto" sulla rete per effettuare le giuste scelte di comunicazione.

Dalla lettura degli articoli su di un dato argomento solitamente ci si aspetta di comprendere più o meno le opinioni su di esso o per lo meno quelle che i *GateKeeper*, coloro che decidono cosa è importante, vogliono far passare.

Nei casi esposti in precedenza si rende necessaria un'analisi delle "Tendenze" di opinione su di un periodo.

Nel lavoro di Wang si prende come obiettivo lo sfruttamento della caratteristica unica del *tweet*, l'*hashtag*.

Su Twitter un *hashtag* è una convenzione spinta dalla comunità per aggiungere contesto e metadati al messaggio.

Gli *hashtag* sono creati dagli utenti come un modo per mettere in luce argomenti e classificare i messaggi.

¹² Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach; Xialong Wang et al.

Questa caratteristica fa percepire Twitter come molto più espressivo degli altri social network. Nel lavoro di Wang si usano 600'000 *tweet*. Solo il 14,6% di questi utilizza almeno un *hashtag*.

Wang classifica gli *hashtag* in tre categorie:

- Topic(#iphone)
- Sentiment(#love)
- Sentiment topic(#iloveobama)

Wang ritiene che il terzo gruppo sia il più espressivo in quanto comprende gli *hashtag* che fondono insieme l'argomento con il giudizio.

L'idea principale di Wang è:

«Aggregare gli hashtag in base alla polarità dei messaggi già classificati in cui compaiono»

Purtroppo questo sistema non dà risultati strabilianti in quanto si basa sulla classificazione automatica dei *tweet* che al giorno d'oggi non è eccezionale. Anzi il principale obiettivo di questo lavoro dovrebbe essere di migliorare le metodologie attuali per fornire nuova informazione, ma se il nuovo modello viene addestrato ad imitare il vecchio ci troviamo nella situazione del cane che si morde la coda.

E' quindi stato utilizzato un secondo metodo, differente da quello originale, per estrarre informazione dagli *hashtag* e le correlazioni tra questi.

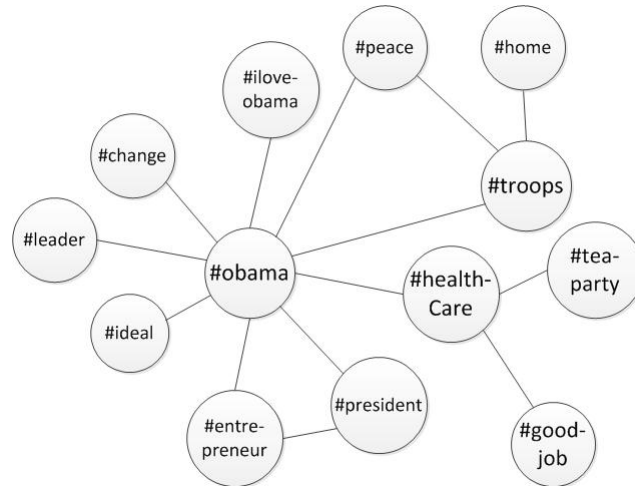
Wang osserva che nel loro *dataset* la probabilità che due *hashtag* della solita polarità ricorrano assieme nel medesimo messaggio è di oltre l'80%. Un altro aspetto che Wang analizza è il significato letterale di un *hashtag*.

Il modello da Wang proposto quindi si basa su un grafico delle co-occorrenze di *hashtag*.

Successivamente Wang costruisce un SVM (*Support Vector Machine*) basato sul significato letterale dello *hashtag*, ma questo non verrà qui analizzato.

La definizione formale dell'obiettivo è dunque:

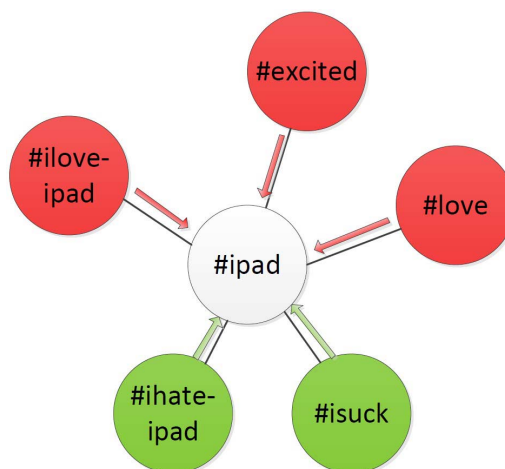
Dato l'insieme di *hashtag* $H=\{h_1,h_2,\dots,h_n\}$ dove ognuno di questi è collegato con un messaggio nell'insieme $T=\{t_1,t_2,t_3,\dots,t_n\}$, classificare la polarità di ogni *hashtag* $Y=\{y_1,y_2,\dots,y_n\}$ attribuendogli un valore contenuto nell'insieme $\{pos,neg\}$.



Dato l'obiettivo e dato il grafo Wang ha cercato di ottenere la polarità di ogni nodo(*hashtag*) basandosi sulla diffusione nella rete; questo concetto verrà approfondito nell'apposito capitolo.

Quindi la polarità di un *hashtag* non è solo figlia dei *tweet* in cui compare ma anche dai suoi vicini nella rete.

Prendendo come esempio #ipad, questo ha come vicini 5 nodi i quali hanno differenti polarità, in verde i negativi e in rosso i positivi.



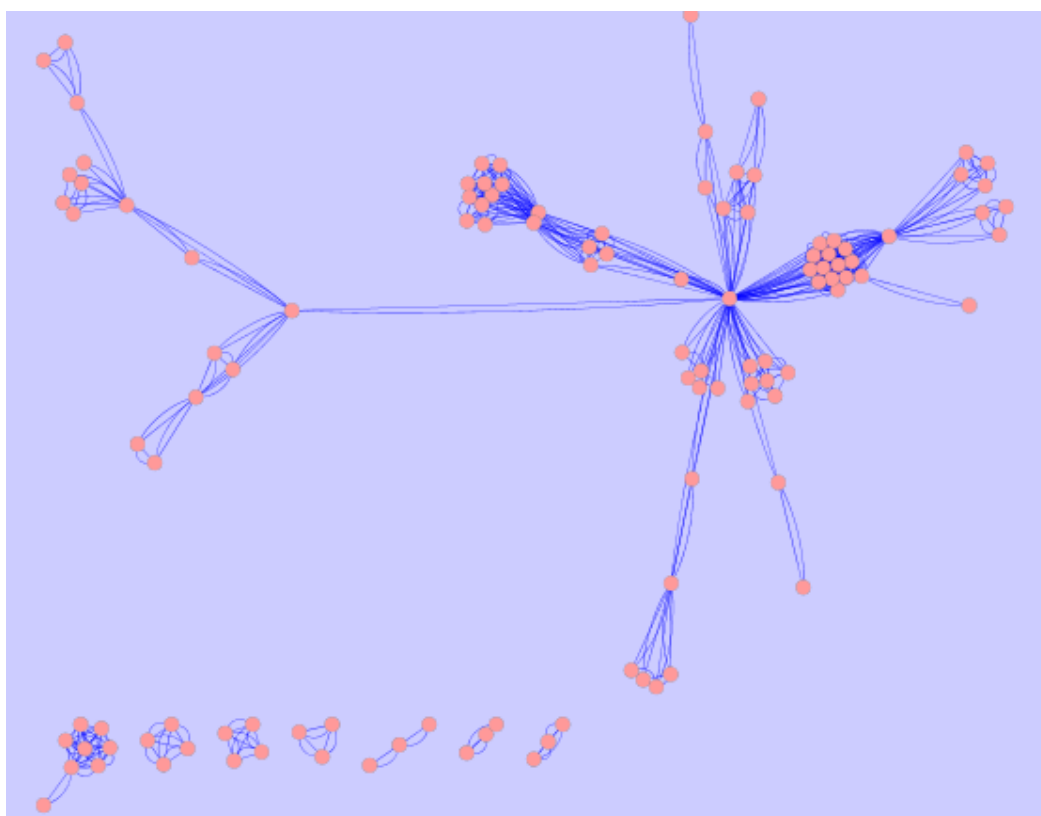
Per raccogliere i dati necessari all'analisi è stato usato un sistema basato sulle *Keyword*. Prima è stata fatta una ricerca sui *topic* d'interesse utilizzando come parole chiavi gli argomenti stessi. Da questi messaggi poi sono state estratte altre parole chiave, in maniera soggettiva supportata da dati lessicali (co occorrenze, ecc.), rilevanti per i vari argomenti.

10. Rete delle Co occorrenze delle *Hashtag*

In questa sezione si analizzerà le varie reti di relazione tra gli *hashtag* seguendo il modello Wang. Inizialmente illustro i dati divisi per mesi, in fine verrà fatta un analisi completa del tutto.

Come già fatto per la rete di *hashtag* per utente i punti d'analisi saranno alcune misurazioni basilari delle reti (Es: grado, centralità) e alcune applicazioni di misure base a nodi ritenuti interessanti. Prendendo per esempio la distanza degli altri nodi dal nodo *Keyword*.

Aprile

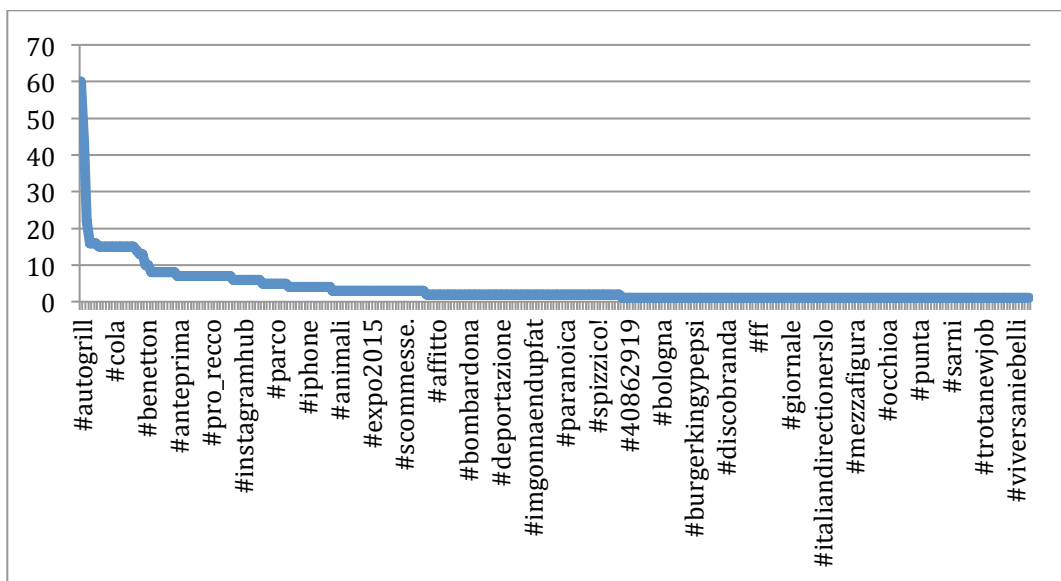


I dati di base

| | |
|----------------------|------|
| Numero di Nodi | 333 |
| Numero di Archi | 1170 |
| Numero di Componenti | 73 |
| Diametro | 7 |

| | |
|--------------------------------------|------------|
| Numero medio di vicini | 3,520 |
| Percorsi più brevi | 30322(27%) |
| Coefficiente di clustering | 0,502 |
| Lunghezza caratteristica dei cammini | 3,254 |

Distribuzione del grado



| I primi 20 | |
|-------------|----|
| #autogrill | 60 |
| #camogli | 43 |
| #liguria | 22 |
| #pallanuoto | 16 |
| #sea | 16 |
| #seriea1 | 16 |
| #25 | 15 |
| #b-day#bea | 15 |
| #beach | 15 |

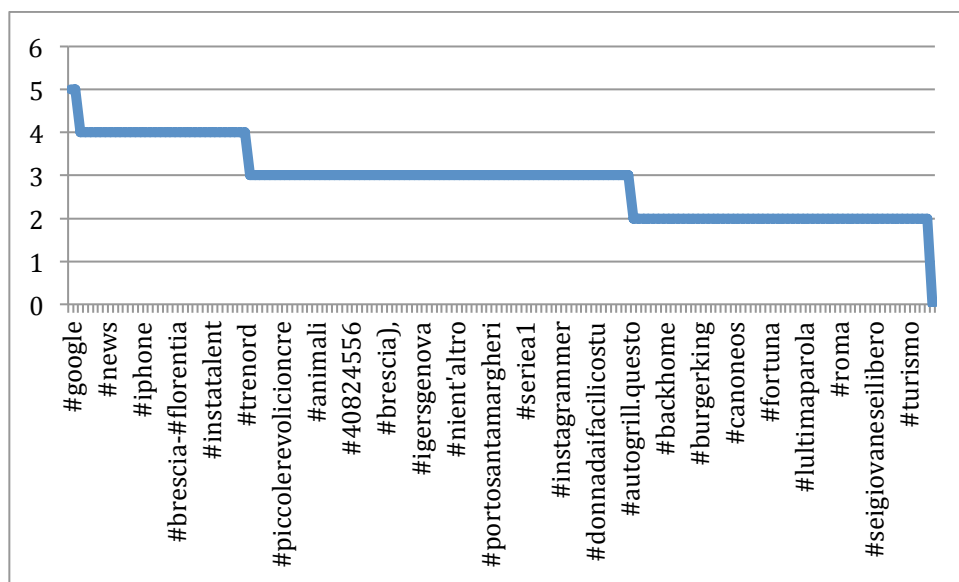
| | |
|-------------|----|
| #belli | 15 |
| #caldo | 15 |
| #cola | 15 |
| #friendship | 15 |
| #ili | 15 |
| #luglio | 15 |
| #me | 15 |
| #s | 15 |
| #sun | 15 |
| #white | 15 |
| #instagram | 14 |

Centralità

| | |
|---------------|------------------|
| #autogrill | 0.206500273924 |
| #camogli | 0.153107441807 |
| #milano | 0.0395819796405 |
| #burgerking | 0.0256660168941 |
| #benetton | 0.0251245397444 |
| #italy | 0.022977742104 |
| #genova | 0.0225127087872 |
| #liguria | 0.0224362649543 |
| #instagram | 0.022321599205 |
| #ultimaparola | 0.0195919173387 |
| #brescia | 0.0141229981271 |
| #sapevatelo | 0.0130336735084 |
| #pallanuoto | 0.00999503115086 |
| #seriea1 | 0.00999503115086 |
| #autogrill. | 0.00825274879282 |
| #animali | 0.00653594771242 |

| | |
|-----------|------------------|
| #news | 0.00653594771242 |
| #notizie | 0.00653594771242 |
| #spizzico | 0.00653594771242 |
| #newipad | 0.00328708481443 |

Distanza da Autogrill



| I 20 più vicini | | I 20 più lontani | |
|-----------------|---|------------------|---|
| #autogrill | 0 | #google | 4 |
| #agrate | 1 | #nu | 4 |
| #anteprima | 1 | #fastfood | 3 |
| #articolo | 1 | #lifestyle | 3 |
| #autogrill! | 1 | #affitto | 3 |
| #autostrade | 1 | #altrisports | 3 |
| #autostrade. | 1 | #casa | 3 |
| #backhome | 1 | #news | 3 |
| #benetton | 1 | #sport | 3 |
| #bombardona | 1 | #40862919 | 3 |

| | | | |
|------------------|---|-------------------------|---|
| #bossi | 1 | #annunci | 3 |
| #breakfast | 1 | #camogli#vintage#barche | 3 |
| #brescia | 1 | #hipstamatic | 3 |
| #buongiorno | 1 | #immobili_terreni | 3 |
| #burgerking | 1 | #iphone | 3 |
| #burgerking, | 1 | #nationalgeographic | 3 |
| #calcioscommesse | 1 | #papaio | 3 |
| #calstorta | 1 | #camogli,sempre | 3 |
| #camogli | 1 | #camogli. | 3 |
| #camogli! | 1 | #waves | 3 |

Autogrill in questo caso si trova vicino agli altri marchi dell'azienda, ad eccezione per #boicotterà. Il collegamento tra #autogrill e #camogli è un ponte in quanto se questo venisse tagliato #Autogrill finirebbe in una componente isolata da quella centrale.

PageRank

I 20 migliori(in ordine crescente)

| | |
|-------------|------------------|
| #luglio | 0.00437121971026 |
| #me | 0.00437121971026 |
| #s | 0.00437121971026 |
| #sun | 0.00437121971026 |
| #white | 0.00437121971026 |
| #autogrill, | 0.00449064551349 |
| #bacon | 0.00449064551349 |
| #sea | 0.00497161231633 |

| | |
|---------------|------------------|
| #italy | 0.00519901190661 |
| #genova | 0.00588868448791 |
| #pallanuoto | 0.00645927142399 |
| #seriea1 | 0.00645927142399 |
| #instagram | 0.00683055359612 |
| #sapevatelo | 0.0070365575392 |
| #burgerking | 0.00703765000275 |
| #milano | 0.00752217413504 |
| #liguria | 0.00757402707731 |
| #benetton | 0.00824245877938 |
| #ultimaparola | 0.00866353301501 |
| #camogli | 0.019248928941 |

I 20 peggiori

| | |
|-----------------------|-------------------|
| #waves | 0.000725655365394 |
| #portosantamargherita | 0.000842040544928 |
| #punta | 0.000842040544928 |
| #garliccheesebread | 0.000876250644196 |
| #40862919 | 0.000903454473588 |
| #fuorisalone | 0.000953372924345 |
| #autogrill.nonostante | 0.00102800023616 |
| #autogrill.questo | 0.00102800023616 |
| #autogrill! | 0.00105357298301 |
| #bossi | 0.00105357298301 |
| #breakfast | 0.00105357298301 |
| #buongiorno | 0.00105357298301 |
| #calcioscommesse, | 0.00105357298301 |
| #camogli! | 0.00105357298301 |
| #carovita | 0.00105357298301 |
| #fortuna | 0.00105357298301 |
| #ignoranzacheamiamo | 0.00105357298301 |

| | |
|---------------|------------------|
| #letteratura | 0.00105357298301 |
| #ultimaparola | 0.00105357298301 |
| #roma | 0.00105357298301 |

Nel mese di aprile vediamo una componente gigante che si sviluppa intorno allo *hashtag* #camogli, si vedrà più avanti che questo particolare nodo tende ad avere una centralità molto alta in tutti i mesi.

La rete è composta da 333 nodi e 1170 archi, quindi possiamo dire che è una rete di medie dimensioni. Il numero di componenti connesse è di 73.

Il coefficiente di *clustering* è medio (0,502), quindi è probabile che la chiusura triangolare si verifichi più o meno nella meta dei casi, si può pensare vedendo questo dato che esistono nodi importanti che si correlano con nodi deboli i quali tendono raramente a connettersi direttamente tra di loro.

La distribuzione del grado è quasi una *Power-Law*, il fatto che non lo sia completamente è una caratteristica delle piccole reti che non riescono a sviluppare a pieno le caratteristiche delle “reti tipiche”.

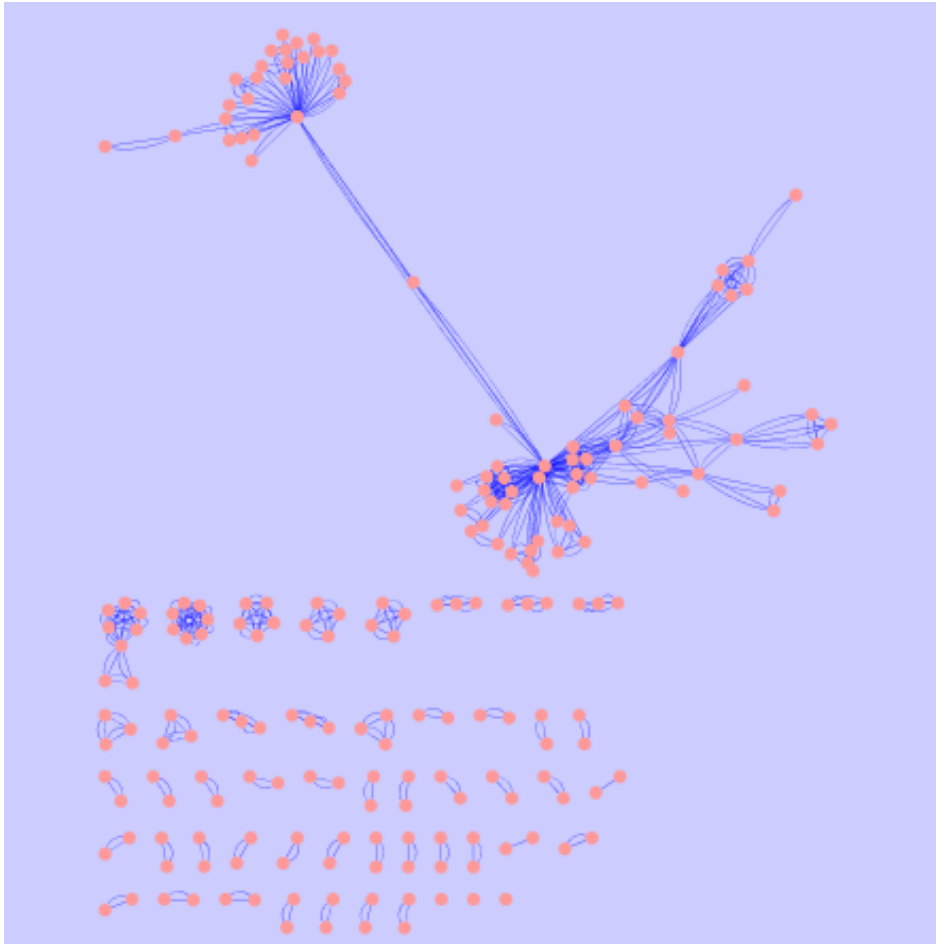
Osservando la classifica per grado si nota che #camogli e #autogrill hanno, ma non si vedono altri *hashtag* collegati direttamente al marchio nei primi 20.

Anche nella classifica per centralità #camogli e #autogrill la fanno da padrone, lasciando #burgerking in quarta posizione, #benetton in quinta che comunque rientrano tra i top 20. Non si vedono altri riferimenti al marchio a parte “#autogrill.”.

Nella classifica per distanza da #Autogrill, troviamo sempre #Brugerking, #Camogli e #benetton.

Il *PageRank* mette in prima posizione #autogrill e in seconda #camogli e #benetton in quarta.

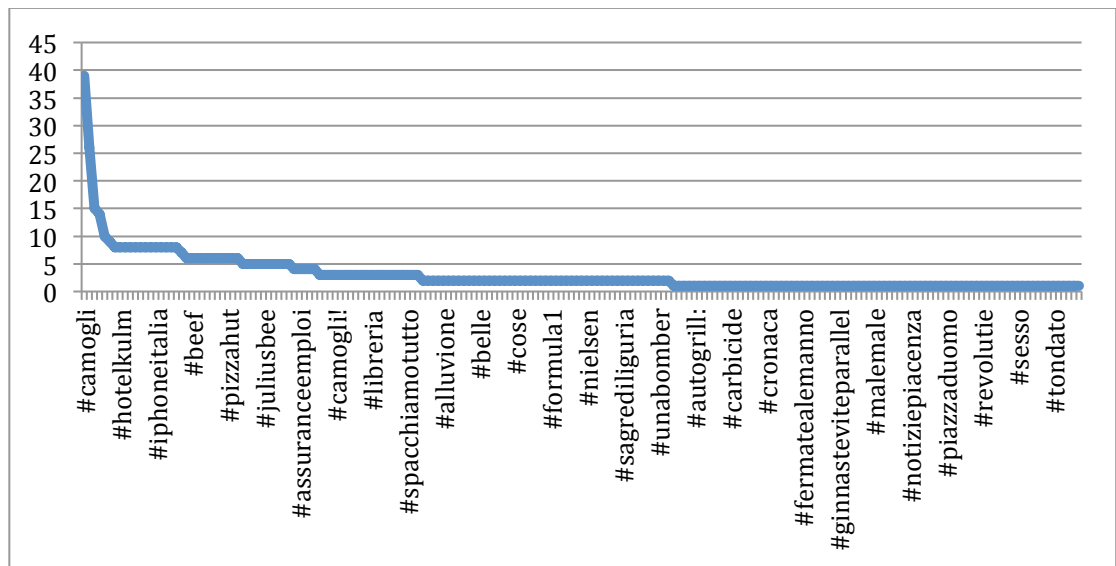
Maggio



I dati di base

| | |
|--------------------------------------|-----------|
| Numero di Nodi | 202 |
| Numero di Archi | 600 |
| Numero di Componenti | 51 |
| Diametro | 6 |
| Numero medio di vicini | 2,980 |
| Percorsi più brevi | 6420(15%) |
| Coefficiente di clustering | 0,532 |
| Lunghezza caratteristica dei cammini | 2,639 |

Distribuzione del grado



I 20 migliori

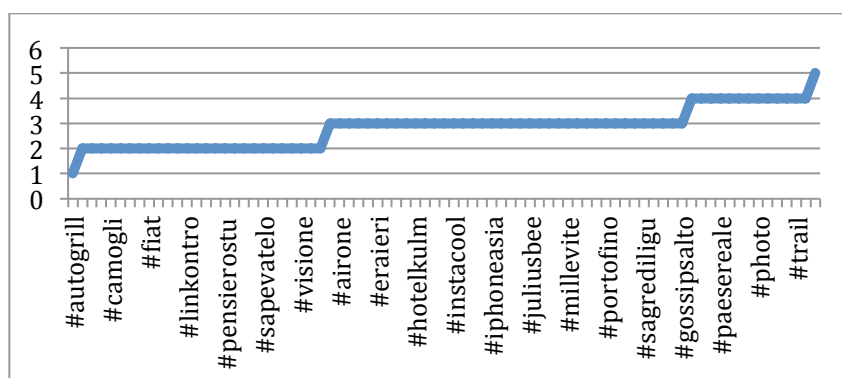
| | |
|----------------|----|
| #camogli | 39 |
| #autogrill | 26 |
| #igers | 15 |
| #italy | 14 |
| #blog | 10 |
| #igersgenova | 9 |
| #genova | 8 |
| #hotelkulm | 8 |
| #ignian | 8 |
| #instacool | 8 |
| #instagood | 8 |
| #instagrammers | 8 |
| #iphon | 8 |
| #iphoneasia | 8 |
| #iphoneitalia | 8 |
| #iphoneography | 8 |
| #popular | 8 |

| | |
|-------------|---|
| #spettacolo | 8 |
| #tramonto | 8 |
| #autogrill. | 7 |

Centralità

| | |
|--------------|-------------------|
| #camogli | 0.124183181098 |
| #autogrill | 0.0863201752043 |
| #blog | 0.0230756904012 |
| #portofino | 0.0120185887506 |
| #italy | 0.00843081744209 |
| #liguria | 0.00827947225041 |
| #genova | 0.00635649805032 |
| #mannarino | 0.00411302815021 |
| #salto12 | 0.00411302815021 |
| #travel | 0.00174492103342 |
| #igers | 0.00162028381675 |
| #autogrill. | 0.000534159500027 |
| #igersgenova | 0.000231469116678 |
| #1 | 00.00 |
| #1mai | 00.00 |
| #2 | 00.00 |
| #3 | 00.00 |
| #5terre | 00.00 |
| #99qc | 00.00 |
| #airone | 00.00 |

Distanza da autogrill



| I 20 più vicini | | I 20 più lontani | |
|-----------------------------|---|------------------|---|
| #autogrill | 0 | #ilfatto | 4 |
| #belle | 1 | #allucinazione | 3 |
| #caffebenecomune | 1 | #crisi | 3 |
| #caffè | 1 | #mannarino | 3 |
| #camogli | 1 | #paesereale | 3 |
| #coffee" | 1 | #terzapagina | 3 |
| #convegnoeste | 1 | #cronaca | 3 |
| #cose | 1 | #china | 3 |
| #fiat | 1 | #photo | 3 |
| #ibs | 1 | #hithotit | 3 |
| #lecco | 1 | #vara | 3 |
| #libreria | 1 | #camogli! | 3 |
| #linkontro | 1 | #trail | 3 |
| #ultimaparola | 1 | #trekking:" | 3 |
| #nielsen | 1 | #5terre | 3 |
| #noncisonopiulemezzstagioni | 1 | #airone | 2 |
| #pensierostupendo:se | 1 | #alluvione | 2 |
| #repubblica | 1 | #appuntamento | 2 |
| #rustichella! | 1 | #blog | 3 |
| #salto12 | 1 | #eraieri | 2 |

PageRank

I top 20(ordine ascendente)

| | |
|----------------|------------------|
| #iphoneography | 0.00592830008504 |
| #popular | 0.00592830008504 |
| #spettacolo | 0.00592830008504 |
| #tramonto | 0.00592830008504 |
| #ignian | 0.00609818837778 |
| #instacool | 0.00609818837778 |
| #instagood | 0.00609818837778 |
| #instagrammers | 0.00609818837778 |
| #iphon | 0.00609818837778 |
| #mannarino | 0.00629539499179 |
| #portofino | 0.00678386251984 |
| #igersgenova | 0.0070237706683 |
| #liguria | 0.00704912628462 |
| #genova | 0.00791015761172 |
| #autogrill. | 0.0079317318749 |
| #blog | 0.00891350553565 |
| #igers | 0.0105570605624 |
| #italy | 0.0114321269107 |
| #camogli | 0.0358591086818 |

I 20 peggiori

| | |
|------------------|------------------|
| #china | 0.00146332417559 |
| #genovapernoi | 0.00155077543614 |
| #lunch | 0.00155077543614 |
| #cronaca | 0.00160968500208 |
| #ilfatto | 0.00166107837571 |
| #airone | 0.00221413154227 |
| #caffebenecomune | 0.00221960705177 |
| #convegnoeste | 0.00221960705177 |
| #ultimaparola | 0.00221960705177 |

| | |
|----------------------------------|------------------|
| #noncisonopiulemezz estagioni | 0.00221960705177 |
| #pensierostupendo:se | 0.00221960705177 |
| #rustichella! | 0.00221960705177 |
| #sapevatelo | 0.00221960705177 |
| #visione | 0.00221960705177 |
| #vitadimmerda | 0.00221960705177 |
| #voglie | 0.00221960705177 |
| #photo | 0.00225514479629 |
| #sagradelpesce | 0.00254940164706 |
| #gossipsalto | 0.00268111756867 |
| #alluvione | 0.0026970007411 |

Nel mese di maggio si presenta una rete di 200 nodi e 600 archi, contenente ben 51 componenti connesse, moltissime delle quali però composte da 2 nodi e che quindi fanno solo rumore. Il coefficiente di *clustering* (0,532) rimane molto vicino a quello del mese precedente (0,561).

La distribuzione del grado mostra una curva più “matura” rispetto al mese precedente.

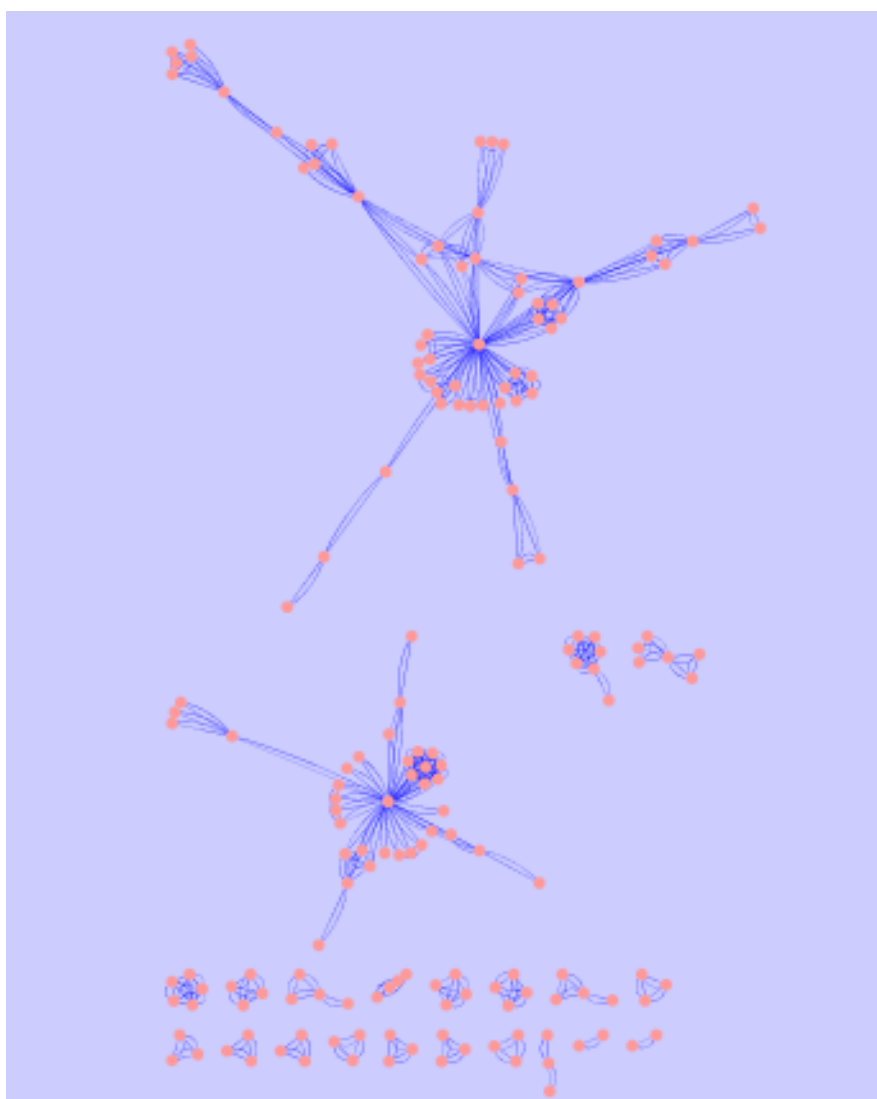
Nella classifica per il grado si trova primo #camogli, secondo #autogrill e un ventesimo #autogrill. .

Nella classifica per centralità ritroviamo primo #camogli, secondo #autogrill e undicesimo #autogrill. .

Nella classifica per distanza da #autogrill troviamo in terza posizione #caffebenecomune, in quarta #caffè, in quinta #camogli e in diciannovesima #rustichella!, ma va comunque segnalato che tutti questi sono a distanza 1 da #autogrill.

Il *PageRank* mette un primo #autogrill e un secondo #camogli.

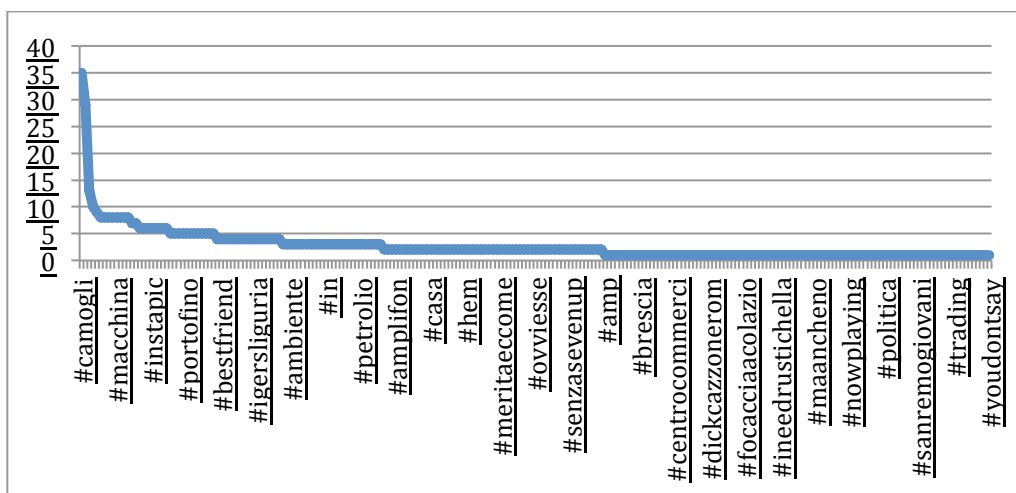
Giugno



I dati di base

| | |
|--------------------------------------|----------|
| Numero di Nodi | 242 |
| Numero di Archi | 644 |
| Numero di Componenti | 60 |
| Diametro | 6 |
| Numero medio di vicini | 2,669 |
| Percorsi più brevi | 5092(8%) |
| Coefficiente di clustering | 0,492 |
| Lunghezza caratteristica dei cammini | 2,55 |

Distribuzione del grado

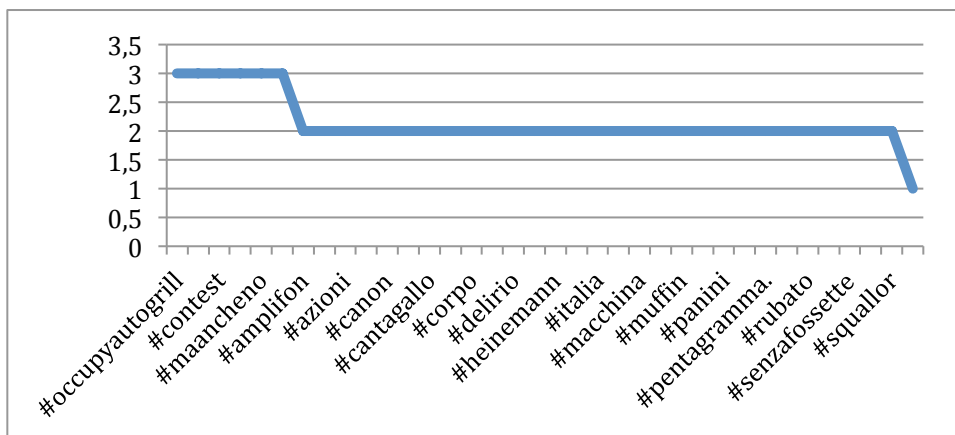


I 20 migliori

| | |
|-----------------|----|
| #camogli | 35 |
| #autogrill | 29 |
| #instagram | 13 |
| #liguria | 10 |
| #igersgenova | 9 |
| #articolo | 8 |
| #canon | 8 |
| #canoneos | 8 |
| #corpo | 8 |
| #macchina | 8 |
| #mia | 8 |
| #obiettivo | 8 |
| #rubato | 8 |
| #camogli, | 7 |
| #genova | 7 |
| #boat | 6 |
| #coloreitaliano | 6 |
| #instacool | 6 |
| #instapic | 6 |
| #photography | 6 |

| Centralità | |
|-----------------|-------------------|
| #camogli | 0.0489588300747 |
| #liguria | 0.0205054840248 |
| #autogrill | 0.0198818825428 |
| #instagram | 0.0119401342577 |
| #camogli, | 0.0101610359121 |
| #genova | 0.00623601481971 |
| #madonna | 0.00418179817321 |
| #pizza | 0.00418179817321 |
| #portofino. | 0.00418179817321 |
| #igersgenova | 0.00393113483242 |
| #sapevatelo | 0.00363156157148 |
| #autogrill, | 0.00212758152672 |
| #ballarò | 0.00124720296394 |
| #italia | 0.00124720296394 |
| #promo | 0.00124720296394 |
| #coloreitaliano | 0.000794786202512 |
| #spizzico | 0.000293459520927 |
| #photography | 0.00018341220058 |
| #euro2012 | 0.000110047320348 |

Distanza da Autogrill



| I 20 più vicini | | I 20 più lontani | |
|-------------------|---|-------------------|---|
| #autogrill | 0 | #occupyautogrill | 2 |
| #amplifon | 1 | #calcio | 2 |
| #articolo | 1 | #contest | 2 |
| #azioni | 1 | #ballaro | 2 |
| #ballarò | 1 | #maancheno | 2 |
| #canon | 1 | #tergicristallo | 2 |
| #canoneos | 1 | #amplifon | 1 |
| #cantagallo | 1 | #articolo | 1 |
| #cinesi; | 1 | #azioni | 1 |
| #corpo | 1 | #ballarò | 1 |
| #crisi: | 1 | #canon | 1 |
| #delirio | 1 | #canoneos | 1 |
| #estate | 1 | #cantagallo | 1 |
| #heinemann | 1 | #cinesi; | 1 |
| #ineedrustichella | 1 | #corpo | 1 |
| #italia | 1 | #crisi: | 1 |
| #italiani | 1 | #delirio | 1 |
| #macchina | 1 | #estate | 1 |
| #mia | 1 | #heinemann | 1 |
| #muffin | 1 | #ineedrustichella | 1 |

PageRank

I top 20 (ordine ascendente)

| | |
|-----------|------------------|
| #articolo | 0.00592362595037 |
| #canon | 0.00592362595037 |
| #mia | 0.00592362595037 |

| | |
|---------------|------------------|
| #rubato | 0.00592362595037 |
| #canoneos | 0.00592362595037 |
| #macchina | 0.00592362595037 |
| #obiettivo | 0.00592362595037 |
| #corpo | 0.00592362595037 |
| #abbazia | 0.00621046652335 |
| #autogrill... | 0.00624231263209 |
| #pizzahut | 0.00624231263209 |
| #sapevatelo | 0.00686622368187 |
| #genova | 0.00730551645573 |
| #igersgenova | 0.00777984348567 |
| #euro2012 | 0.00816561389777 |
| #spizzico | 0.00869387418065 |
| #camogli, | 0.00888316585898 |
| #liguria | 0.010141689709 |
| #instagram | 0.0103212671989 |
| #autogrill | 0.0300109494657 |
| #camogli | 0.0322990543745 |

I 20 peggiori

| | |
|-----------------|------------------|
| #habawaba | 0.00142270347857 |
| #liguria: | 0.00142270347857 |
| #teatrosociale | 0.00142270347857 |
| #waugh | 0.00142270347857 |
| #photo | 0.0014479446716 |
| #camogli! | 0.00150034149762 |
| #cantagallo | 0.00151792914978 |
| #delirio | 0.00151792914978 |
| #heinemann | 0.00151792914978 |
| #ineedrusicella | 0.00151792914978 |
| #panino?! | 0.00151792914978 |
| #pentagramma. | 0.00151792914978 |
| #politica | 0.0015253962991 |

| | |
|------------------|------------------|
| #occupyautogrill | 0.00159235769822 |
| #dukan.. | 0.00171696801242 |
| #mordiefuggi | 0.00171696801242 |
| #contest | 0.00183843454315 |
| #calcio | 0.00183843454315 |
| #ballaro | 0.00209737040473 |
| #maancheno | 0.00209737040473 |

A giugno si crea una rete di 242 nodi e 644 archi, un po' più grande delle precedenti. Il numero di componenti è di 60 e il coefficiente di *clustering* si ferma a 0,492, è sceso abbastanza.

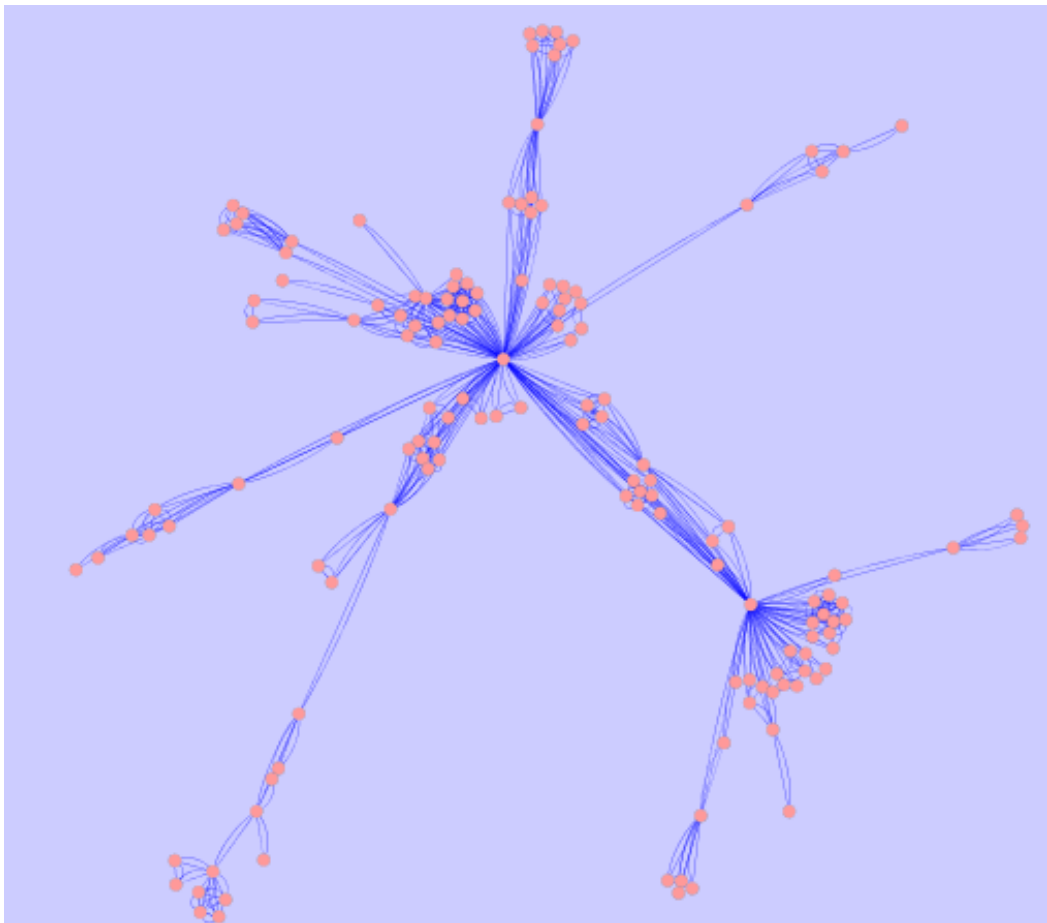
Nella classifica per grado troviamo sempre ai primi posti #camogli(1°), #autogrill(2°).

#camogli rimane in testa alla classifica anche per centralità mentre #autogrill scende al terzo posto e si presenta uno spizzico in diciassettesima posizione.

In base alla distanza da #autogrill troviamo a distanza 1 troviamo diversi *hashtag* tra i quali #amplifon, #azioni, #cantagallo, #ineedrustichella, #muffin.

Il *PageRank* da il primo posto a #camogli e il secondo ad #autogrill. Nella lista dei peggiori troviamo un #occupyautogrill con un valore di *PageRank* bassissimo.

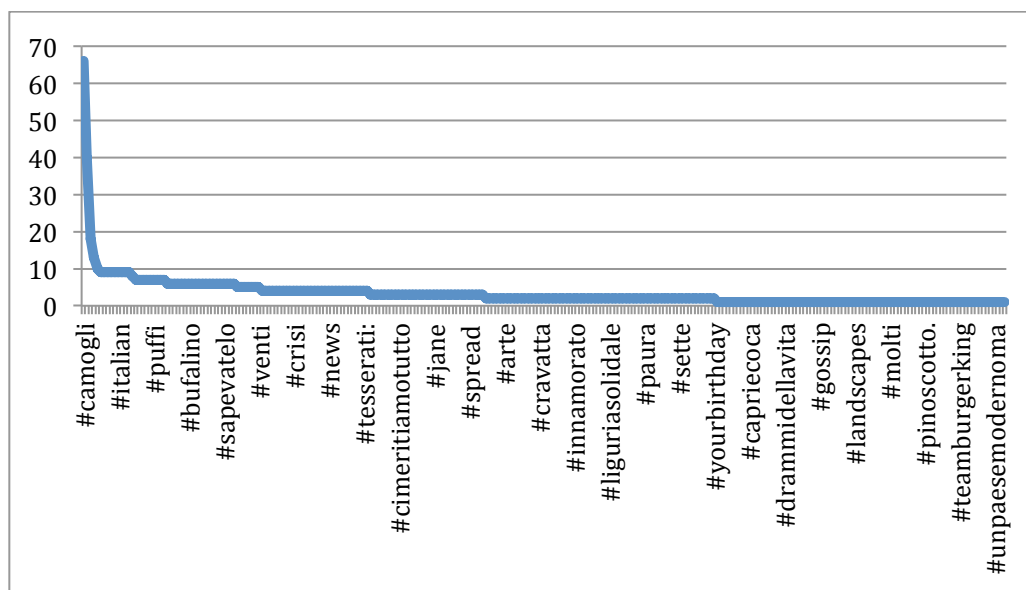
Luglio



I dati di base

| | |
|--------------------------------------|------------|
| Numero di Nodi | 269 |
| Numero di Archi | 892 |
| Numero di Componenti | 50 |
| Diametro | 8 |
| Numero medio di vicini | 3,316 |
| Percorsi più brevi | 21144(29%) |
| Coefficiente di clustering | 0,601 |
| Lunghezza caratteristica dei cammini | 3,255 |

Distribuzione del grado



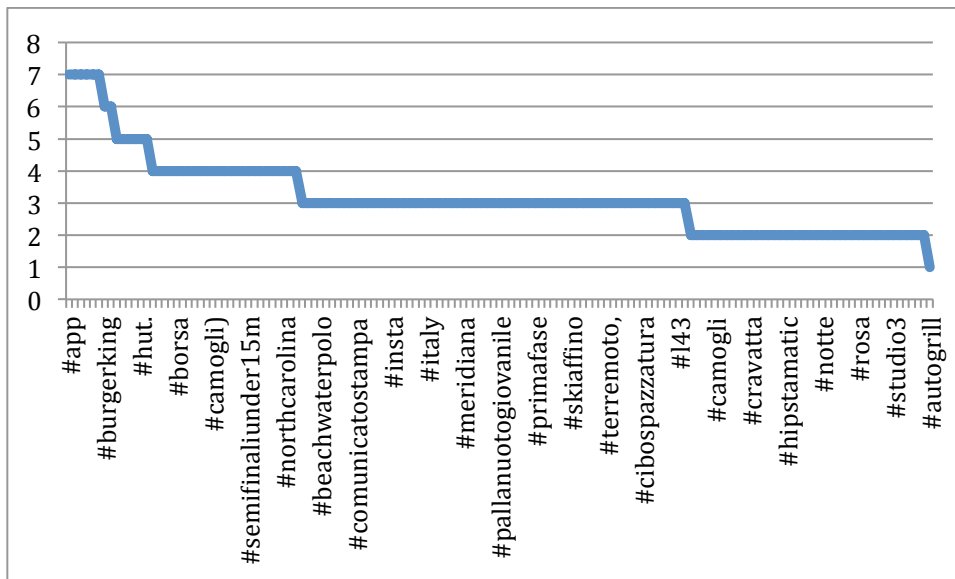
I top 20

| | |
|----------------------|----|
| #camogli | 66 |
| #autogrill | 40 |
| #liguria | 18 |
| #pallanuotogiovanile | 13 |
| #genova | 10 |
| #cavezzo | 9 |
| #insta | 9 |
| #instag | 9 |
| #instagood | 9 |
| #iphone4s | 9 |
| #italian | 9 |
| #landscape | 9 |
| #niceview | 9 |
| #pool | 9 |
| #rosa | 8 |
| #andreavetralla | 7 |
| #burgerking | 7 |
| #facebook | 7 |
| #modena | 7 |
| #mt | 7 |

Centralità

| | |
|------------------|------------------|
| #camogli | 0.248604378912 |
| #autogrill | 0.134564962896 |
| #genova | 0.0535802397469 |
| #news | 0.0427538966128 |
| #fail | 0.0317824282356 |
| #pallanuotogiova | |
| nile | 0.0283284474501 |
| #burgerking | 0.0242649406438 |
| #usa | 0.0240327402548 |
| #sapevatelo | 0.0163991524686 |
| #italia | 0.0162540272255 |
| #saratommasi | 0.0123646707108 |
| #mare | 0.00824311380722 |
| #beachwaterpolo | 0.00811250108844 |
| #rivieracup2012 | 0.00811250108844 |
| #pizza | 0.00619684787972 |
| #liguria | 0.00595497247458 |
| #rosa | 0.00511808357279 |
| #economia | 0.00415058195222 |
| #italy | 0.00415058195222 |
| #ontheroad | 0.00415058195222 |

Distanza



I 20 più vicini

| I 20 più vicini | | I 20 più lontani | |
|------------------|---|--------------------|---|
| #autogrill | 0 | #app | 6 |
| #aelia | 1 | #arte | 6 |
| #andreavetralla | 1 | #bkmobilecrowncard | 6 |
| #apollo | 1 | #damienhirst | 6 |
| #bufalino | 1 | #fastfood | 6 |
| #camogli | 1 | #payment: | 6 |
| #cantagallo | 1 | #burgerking | 5 |
| #capri. | 1 | #pastone | 5 |
| #carpark | 1 | #fail | 4 |
| #chianti | 1 | #fired | 4 |
| #corsera). | 1 | #firenze | 4 |
| #cravatta | 1 | #ilg | 4 |
| #drammidellavita | 1 | #hut. | 4 |
| #dufry | 1 | #bacon | 4 |
| #dutyfree | 1 | #affitto | 3 |
| #facebook | 1 | #casa | 3 |
| #granduca | 1 | #news | 3 |
| #hipstamatic | 1 | #autogrill: | 3 |
| #inas1982 | 1 | #borsa | 3 |
| #jane | 1 | #economia | 3 |

PageRank

I top 20 (ordine decrescente)

| | |
|------------|------------------|
| #pool | 0.0050291165123 |
| #instagood | 0.0050291165123 |
| #italian | 0.0050291165123 |
| #niceview | 0.0050291165123 |
| #usa | 0.00512255802201 |
| #crisi | 0.00537984647623 |

| | |
|----------------------|------------------|
| #non | 0.00552825445795 |
| #time | 0.00552825445795 |
| #camogli. | 0.0055566040475 |
| #pizza | 0.00571289025083 |
| #rosa | 0.00584764175126 |
| #saratommasi | 0.00591713560622 |
| #cavezzo | 0.00624482378759 |
| #autogrill. | 0.0067131031698 |
| #burgerking | 0.00706035579131 |
| #sapevatelo | 0.00736218961736 |
| #genova | 0.00771408200702 |
| #pallanuotogiovanile | 0.0107762191992 |
| #liguria | 0.0112356662272 |
| #autogrill | 0.0326466962351 |
| #camogli | 0.0470197015116 |

I 20 peggiori

| | |
|---------------------------|------------------|
| #ff | 0.00109875494552 |
| #camping | 0.00117373858003 |
| #ilsecoloxix | 0.00117373858003 |
| #nonvogliotornareincitta' | 0.00117373858003 |
| #openliguria366 | 0.00117373858003 |
| #panorama | 0.00117373858003 |
| #snapseed | 0.00117373858003 |
| #chianti | 0.00126192411317 |
| #drammidellavita | 0.00126192411317 |
| #novara | 0.00126192411317 |
| #unpaesemodernomai! | 0.00126192411317 |
| #versobrunico | 0.00126192411317 |
| #gopro | 0.00131559061646 |
| #hut. | 0.00137750793703 |
| #cibospazzatura | 0.00146296611599 |
| #capriecoca | 0.00151920476724 |
| #milano | 0.00151920476724 |
| #ilg | 0.0015666424965 |
| #firenze | 0.00156897890721 |
| #pastone | 0.00161050214309 |

Nel mese di luglio la rete degli *hashtag* è scesa a 269 nodi e 892 archi, con un numero di componenti pari a 50. Il coefficiente di *clustering* è di 0,601, vicino al mese precedente.

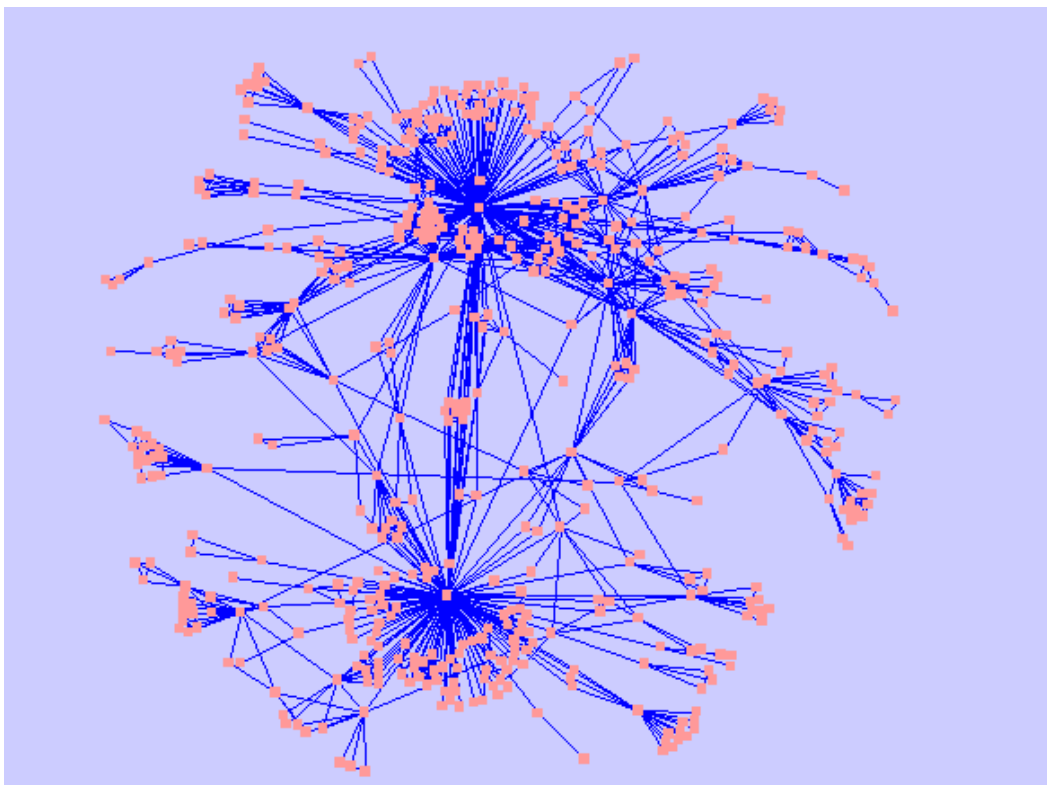
La distribuzione dei gradi mostra una rete abbastanza matura, #camogli raggiunge addirittura i 66 archi, cosa che non si era mai verificata fino ad ora, mentre #autogrill si attesta in seconda posizione con 40 archi.

Per la centralità, #camogli guida ancora la classifica distaccando #burgerking in settima posizione e in nona posizione si trova #sapevatelo. #autogrill è in seconda posizione.

Tra i più vicini ad #autogrill troviamo a distanza 1 : #apollo, #bufalino, #camogli, #carpark e #dutyfree.

Per il *PageRank* il primo posto va come al solito a #camogli e il secondo ad #autogrill, un sesto posto a #sapevatelo e un settimo a #burgerking.

Totale



I dati di base

| | |
|--------------------------------------|-------------|
| Numero di Nodi | 873 |
| Numero di Archi | 3074 |
| Numero di Componenti | 152 |
| Diametro | 8 |
| Numero medio di vicini | 3,523 |
| Percorsi più brevi | 263292(34%) |
| Coefficiente di clustering | 0,529 |
| Lunghezza caratteristica dei cammini | 3,473 |

Grado

I top 20

| | |
|--------------|-----|
| #camogli | 159 |
| #autogrill | 135 |
| #liguria | 50 |
| #instagram | 29 |
| #italy | 28 |
| #genova | 23 |
| #sea | 21 |
| #autogrill. | 17 |
| #igersgenova | 17 |
| #pallanuoto | 17 |
| #burgerking | 16 |
| #caldo | 16 |
| #instagood | 16 |
| #me | 16 |
| #milano | 16 |

| | |
|------------|----|
| #portofino | 16 |
| #seriea1 | 16 |
| #25 | 15 |
| #b-day#bea | 15 |
| #beach | 15 |

Centralità

I 20 più centrali

| | |
|-------------|------------------|
| #camogli | 0.222243783158 |
| #autogrill | 0.20255561222 |
| #burgerking | 0.0224380919833 |
| #milano | 0.0216925596043 |
| #liguria | 0.0215257576022 |
| #usa | 0.0212746075807 |
| #instagram | 0.0200748494154 |
| #genova | 0.0183593508556 |
| #portofino | 0.016101051869 |
| #italy | 0.0156418555017 |
| #autogrill. | 0.0151070203532 |
| #sapevatelo | 0.0149296167546 |
| #blog | 0.0147820908974 |
| #spizzico | 0.0131479864085 |
| #pizzahut | 0.0129563232834 |
| #news | 0.0112569576725 |
| #pizza | 0.0104351211876 |
| #camogli | 0.00989489250039 |
| #autogrill | 0.00952748809064 |
| #burgerking | 0.00934758653855 |

Distanza da Autogrill

| I 20 più vicini | | I 20 più distanti | |
|-----------------|---|-------------------|---|
| #autogrill | 0 | #mayapensatecivoi | 6 |
| #aelia | 1 | #sanroccodi... | 6 |
| #agrate | 1 | #serviceexport | 6 |
| #amplifon | 1 | #barbeque | 6 |
| #andreavetralla | 1 | #nashville | 6 |
| #anteprima | 1 | #landscapes" | 6 |
| #apollo | 1 | #camogli: | 5 |
| #articolo | 1 | #cimeritiamotutto | 5 |
| #autogrill! | 1 | #spread | 5 |
| #autostrade | 1 | #squinzometro | 5 |
| #autostrade. | 1 | #tristecamogli | 5 |
| #azioni | 1 | #ilfatto | 5 |
| #backhome | 1 | #firenze | 5 |
| #belle | 1 | #google | 5 |
| #benetton | 1 | #lavoro | 5 |
| #bombardona | 1 | #nu | 5 |
| #bossi | 1 | #offertelavoro | 5 |
| #breakfast | 1 | #bacon | 5 |
| #brescia | 1 | #hut | 5 |
| #bufalino | 1 | #hut. | 5 |

Pagerank

I 20 migliori (ordine ascendente)

| | |
|--------------|------------------|
| #igersgenova | 0.00280082534909 |
| #sea | 0.00281788642697 |
| #camogli, | 0.00283554854954 |
| #benetton | 0.00284782484399 |

| | |
|----------------------|------------------|
| #pallanuoto | 0.00287167080529 |
| #ultimaparola | 0.0031506402913 |
| #pallanuotogiovanile | 0.00318059262907 |
| #pizzahut | 0.0032533605864 |
| #pizza | 0.00326889648233 |
| #portofino | 0.00335348692695 |
| #milano | 0.00345999749992 |
| #spizzico | 0.00407936598546 |
| #autogrill. | 0.00419193649267 |
| #genova | 0.00468194392002 |
| #burgerking | 0.00493535269782 |
| #instagram | 0.00506846330565 |
| #italy | 0.00511002930494 |
| #sapevatelo | 0.00598032709091 |
| #liguria | 0.00796647232068 |
| #camogli | 0.030617177389 |
| #autogrill | 0.0363962068965 |

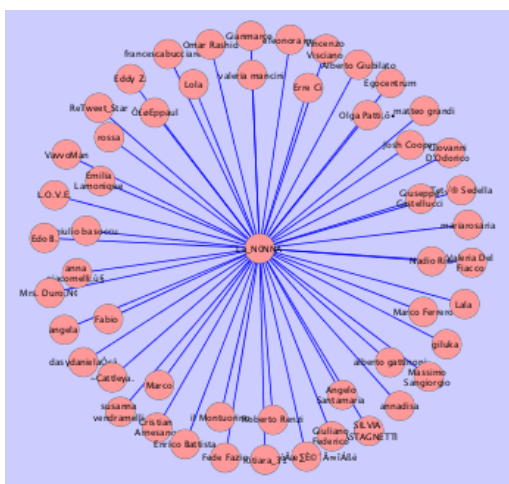
In questo tipo di rete (*hashtag - hashtag*) #camogli supera d'importanza #autogrill, rispetto alla rete del capitolo 8. Si nota nella classifica per grado la presenza di #burgerking e addirittura per centralità arriva ad 3°, quindi è una componente fondamentale per la struttura della rete. #spizzico invece si trova in 13° posizione per centralità mentre invece #sapevatelo (critica) si trova in 11° . Tra gli *hashtag* più vicino ad #autogrill si trovano #apollo, #benetton, #bufalino e #autostrade. Non si rilevano *hashtag* negativi vicino ad #autogrill e questa è una informazione positiva.

11. Cosa si dice su Twitter a Giugno

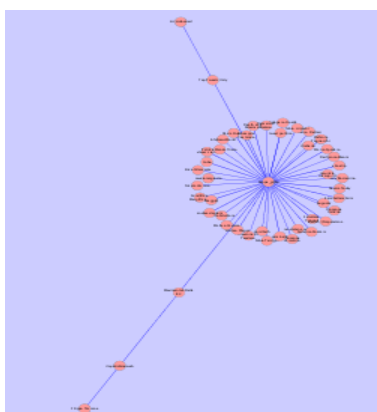
Questa sezione mostra un'analisi su diverse parole chiave, scelte in base all'attività durante il mese. Questo è un modello di analisi preliminare volta a riassumere cosa è accaduto nel periodo, se venisse riscontrato qualche fenomeno interessante, dovrebbe essere approfondito con un altro lavoro.

Relazioni di Retweet interessanti:

Autogrill:



Questo gruppo di persone ha retwittato un post satirico dell'utente **LA_NONNA**. Una cosa molto simile accade con l'utente **matteograndi**, il quale fa una satira dicendo che in autogrill si festeggia il compleanno di Fabio Volo.



Qui a fianco invece è mostrata la diffusione di un messaggio appartenente a **beppe_grillo**, il quale ironizza sul fatto che l'autostrada Reggio-Calabria può essere definita tale solo per la presenza degli Autogrill. Bisogna notare in questo grafico che persone che hanno retwittato, sono state retwittate a loro volta.

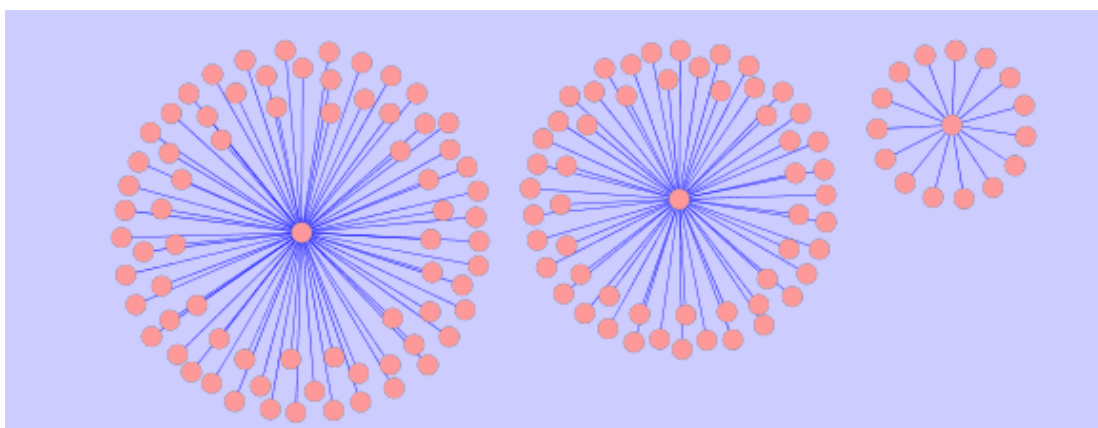
Oltre la struttura della rete, molti messaggi hanno come oggetto I seguenti argomenti:

- Prezzi un po' alti;
- Pizza idroponica in Autogrill(non si capisce se è una Battuta o una richiesta);

- Condizione igieniche dei bagni Autogrill;
- Divise dell'Autogrill simili a quelle della Croazia.

Tirando le somme non si riscontrano criticità degne di nota in questi messaggi, per lo meno nei più diffusi. La satira non offensiva aiuta a fare in modo che si parli del marchi nel gergo comune, anche se non direttamente.

Burger King:

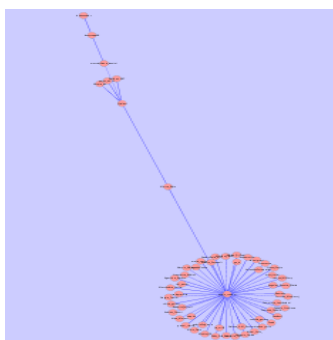


Anche del marchio Burger king si parla principalmente in maniera satirico/giocosa non mostrando nessun attacco al marchio, e questo è decisamente buono pensando ad altri marchi su internet.

Spizzico:

Per Spizzico si nota che l'utente **fabiovoloo** (probabilmente non è il Vip) ironizza che non è possibile andare da Spizzico perché il papa ha prenotato tutti i tavoli.

Culatello:



Nell'immagine accanto si vede una comunità di persone che vuole una maggiore diffusione dei prodotti emiliani. Il tutto è partito da un messaggio di **paola_saluzzi**, la quale dice di essere stanca delle solite diete e di voler mangiare prodotti emiliani.

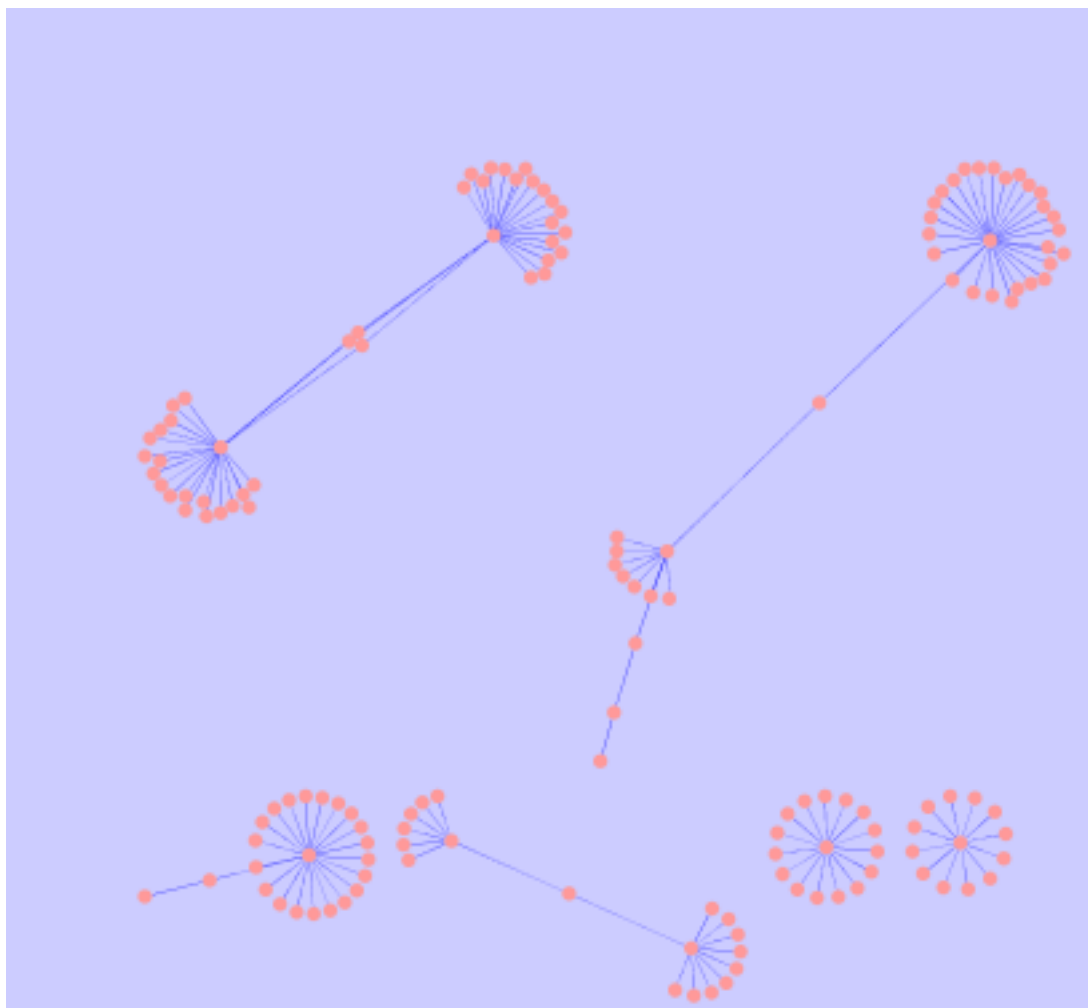
Camogli:

In questa rete a livello di struttura non ci sono cose rilevanti. A livello del singolo messaggio c'è un messaggio indirizzato all'azienda da parte di **mocchina:**

«Signor Benetton: li cambiamo un po' sti panini o vogliam tirare avanti altri 10 anni con Camogli e Capri? #Autogrill»

12.Retweet a Luglio

In questa sezione si analizzerà la rete di *retweet* limitata esclusivamente alla parola chiave “autogrill”. L’analisi è ideata come un report riassuntivo del mese, come già visto per giugno e non come analisi approfondita.



Le chiacchiere su Sara Tommasi che scompare in un autogrill la fanno da padrone. Le prime tre principali componenti parlano di questo con i seguenti messaggi:

«Jean Ralphio La timidezza gioca veramente brutti scherzi, anche a Sara "Glory Hole" Tommasi <http://t.co/tQJRFCsj> Mon, 02 Jul 2012 »

«RT @el_buddy: Ansa: Sara Tommasi diserta il debutto nel night e scompare in autogrill. Ritrovata a Parma sotto Shock, giovane camionista Rumeno. Mon, 02 Jul 2012 »

«RT @articolisti: Sara Tommasi scompare per ore in autogrill. Stava aggiornando gli annunci in bagno. Mon, 02 Jul 2012 »

«RT @marcosalvati: Sara Tommasi non si era persa in autogrill, si stava solo facendo un Camogli. Lo conferma anche Giorgio Camogli. Mon, 02 Jul 2012”

«RT @AndyViolet: Tommasi scompare in autogrill a Lodi, appare a Parma. Che fa nel tragitto? SARA TOMMASI IN AUTOGRILL = ***** SUL TRAM #anagrammi Mon, 02 Jul 2012 »

Un messaggio che si discosta di molto dal gruppo è il seguente:

«katioursillo autogrill RT @espressonline: Per le donne non c'è lavoro - Red Bull, Rai, Omsa, Autogrill: i più recenti casi di discriminazione sul posto di... <http://t.co/DHsHUUhl> Fri, 13 Jul 2012»

«kong fusion autogrill RT @THEREALGUE: Capolavori all'autogrill <http://t.co/8wKDYz8Q> Wed, 04 Jul 2012 »

Il primo è importante per l'argomento trattato, si parla di discriminazione delle donne nell'assunzioni di Autogrill.

Il secondo invece è un commento al fatto che in Autogrill si vendano copie del film “Io vi troverò” ritenuto dall'utente un capolavoro. In questo caso l'utente usa Instangram.com che è un servizio di condivisione fotografie. In questo caso l'informazione non era direttamente nel messaggio, ma vi si trova solo un riferimento e questo comportamento è abbastanza diffuso su Twitter.

13. Proteggere i propri dati in un sistema per condividere i propri dati

Sembra illogico parlare di *privacy* su di un *social network*, ma invece è uno degli aspetti cruciali per la comprensione di come il dato sia creato, acquisito e diffuso su queste piattaforme e al di fuori di esso. Qui di seguito verranno riassunte le politiche dei principali *social network* esistenti.

Facebook e la sua politica

Iniziare l'analisi da Facebook è più che logico, è dovuto. E' il più grande *social network* esistente e quello che sta al centro di una nuova visione della *privacy*,

Lo stesso creatore di Facebook ha affermato durante un'intervista che le nuove generazioni stanno rivedendo il concetto di *privacy*. «Finita l'era della *privacy*», così scrive Nicola Bruno, sul *corriere della sera*, citando Mark Zuckerberg¹³.

«Ormai gli utenti condividono senza problemi le informazioni personali online. Le norme sociali cambiano nel tempo. E così è anche per la privacy»

Zuckerberg

Secondo il fondatore di Facebook, l'era della condivisione online è iniziata con l'avvento dei blog personali, dove ogni persona parla dei propri interessi ed esprime la propria opinione. Facebook ha permesso di condividere le informazioni degli utenti con chi si desidera all'interno della rete. Ha fatto molto discutere il cambiamento della politica sulla *privacy* rispetto alle sue origini, all'inizio le informazioni pubbliche rimanevano comunque nel *social network*, mentre dal Gennaio del 2011 sono rintracciabili anche dai motori di ricerca esterni.

¹³ http://www.corriere.it/scienze_e_tecnologie/10_gennaio_11/finita-era-privacy-internet_e8a18cf0-febe-11de-a5d5-00144f02aabe.shtml

Il problema è che molti utenti non si sono resi conto di quanto fosse realmente pubblico il contenuto reso pubblico, ovvero si aspettavano che i dati anche se visibili da altri rimanessero comunque di loro proprietà e non di tutti. La Francia ha portato avanti un progetto di legge per garantire il *diritto all'oblio dell'informazione*, il quale metterebbe una scadenza alle informazioni condivise online. Il diritto all'oblio è tecnicamente, in molte parti del mondo, utilizzato solo per gli atti giudiziari e non per le informazioni personali in generale.

«È riconosciuto un "diritto all'oblio", cioè il diritto a non restare indeterminatamente esposti ai danni ulteriori che la reiterata pubblicazione di una notizia può arrecare all'onore e alla reputazione, salvo che, per eventi sopravvenuti, il fatto precedente ritorni di attualità e rinasca un nuovo interesse pubblico all'informazione. Analogo principio è stato applicato anche a personaggi che hanno avuto grande notorietà.»

Repubblica.it

Per comprendere meglio il problema è necessario leggere che cosa dice la regolamentazione di Facebook andando nella sezione apposita del *Social Network*.

Come Facebook spiega la sua politica

Quanto segue è preso direttamente dalla sezione *privacy* di Facebook

Informazioni ricevute da Facebook sugli utenti

Facebook riceve diversi tipi di informazioni su di te, tra cui:

Informazioni personali

Le tue informazioni includono quelle richieste per l'iscrizione al sito e qualsiasi altra informazione che decidi di condividere.

Informazioni di registrazione

Quando ti iscrivi a Facebook, ti viene richiesto di fornire il nome, l'indirizzo email, la data di nascita e il sesso.

Informazioni che decidi di condividere

Le tue informazioni includono anche le informazioni che condividi intenzionalmente su Facebook, ad esempio quando pubblichi un aggiornamento dello stato, carichi una foto o commenti la notizia di un amico.

Sono incluse anche le informazioni che condividi intenzionalmente quando intraprendi un'azione, ad esempio se aggiungi un amico, se ti piace una Pagina o un sito Web, se aggiungi un luogo in una delle tue notizie, se trovi gli amici usando il nostro strumento per l'importazione dei contatti o se indichi la tua situazione sentimentale.

Il tuo nome, le immagini del profilo, le foto di copertina, le reti, il nome utente e l'ID utente vengono trattati come le altre informazioni che decidi di rendere pubbliche.

Conoscendo la tua data di nascita, Facebook può ad esempio mostrarti contenuti e pubblicità adatti alla tua età.

Informazioni sugli utenti condivise dagli altri

Riceviamo informazioni su di te dai tuoi amici e da altre persone, ad esempio quando caricano le tue informazioni di contatto, pubblicano una tua foto, ti taggano in una foto, in un aggiornamento di stato o in un luogo o ti aggiungono a un gruppo.

Quando le persone utilizzano Facebook, possono memorizzare e condividere le informazioni di cui dispongono su di te e su altre persone, ad esempio quando caricano e gestiscono i loro inviti e contatti.

(Questo andrebbe approfondito, gli altri condividono informazioni su di te)

Altre informazioni ricevute da Facebook sugli utenti

Facebook riceve anche altri tipi di informazioni sugli utenti:

- Facebook riceve i tuoi dati personali ogni volta che interagisci con il sito, ad esempio quando visiti il diario di un'altra persona, invii o ricevi un messaggio, cerchi un amico o una Pagina, clicchi, visualizzi o interagisci in altro modo con gli elementi, usi un'applicazione

Facebook mobile, acquisti crediti Facebook o acquisti articoli tramite Facebook.

- Quando pubblichi elementi come foto o video, Facebook potrebbe ricevere ulteriori dati correlati (o dati multimediali), come ad esempio la data, l'ora e il luogo in cui hai scattato la foto o girato il video.

- Riceviamo dati dal computer, dal cellulare o dagli altri dispositivi che utilizzi per accedere a Facebook e anche quando più utenti effettuano l'accesso dallo stesso dispositivo. Tali dati possono includere l'indirizzo IP e altre informazioni su cose come il tuo servizio Internet, il luogo, il tipo di browser in uso (compresi i dati di identificazione) o le Pagine che visiti. Ad esempio, possiamo ottenere la tua posizione GPS o altre informazioni di posizione, in modo da poterti informare se ci sono amici nelle vicinanze.

- Quando visiti un gioco, un'applicazione o un sito Web che usa la Piattaforma Facebook oppure visiti un sito con una funzione di Facebook (come un plug-in sociale), riceviamo delle informazioni, a volte tramite i cookie, che possono includere la data e l'ora in cui hai visitato il sito, l'indirizzo Web, o URL, su cui ti trovi, delle informazioni tecniche sull'indirizzo IP, il browser e il sistema operativo che usi, se hai effettuato l'accesso a Facebook, e il tuo ID utente.

- Talvolta, Facebook riceve dati dai propri inserzionisti partner, clienti e altre terze parti al fine di migliorare la qualità delle inserzioni, aiutare gli inserzionisti stessi a comprendere l'attività online e migliorare Facebook in generale. Ad esempio, un inserzionista potrebbe comunicare a Facebook informazioni su di te (come la tua risposta a un'inserzione su Facebook o su un altro sito) per valutare l'efficacia e migliorare la qualità di tali inserzioni.

Facebook combina inoltre i dati che ha a disposizione dalle informazioni su di te e sui tuoi amici. Ad esempio, potremmo raccogliere dati su di te per determinare quali amici mostrarti nella sezione Notizie o suggerirti tag nelle

foto che pubblichi. Potremmo associare la tua città attuale a informazioni GPS e di altro tipo che abbiamo a disposizione sulla tua posizione, ad esempio per far sapere a te e ai tuoi amici della presenza di eventi nelle vicinanze che potrebbero interessarvi. Potremmo anche aggregare dati su di te per mostrarti inserzioni più interessanti.

Quando risaliamo alla tua posizione GPS, la combiniamo con altre informazioni sulla posizione che abbiamo a disposizione su di te (come la tua città attuale). Tuttavia, manteniamo tali informazioni solo finché sono utili per fornirti i nostri servizi, ad esempio manteniamo le tue ultime coordinate GPS per poterti inviare notifiche pertinenti.

Forniamo dati ai nostri inserzionisti partner o clienti solo dopo aver rimosso il tuo nome e qualsiasi altra informazione personale o dopo aver combinato i tuoi dati con quelli di altre persone in modo che non siano più associati a te.

Informazioni pubbliche

Con l'espressione "Informazioni pubbliche" (o "Informazioni visibili a tutti"), si intendono le informazioni che decidi di rendere pubbliche, nonché le informazioni sempre disponibili pubblicamente.

Informazioni che decidi di rendere pubbliche

Se decidi di rendere pubbliche le tue informazioni, **tutti**, comprese le persone al di fuori di Facebook, potranno vederle.

Scegliere di rendere pubbliche le proprie informazioni significa anche che tali informazioni:

- potranno essere associate a te (ad es. nome, immagini del profilo, foto di copertina, diario, ID utente, nome utente, ecc.) anche al di fuori di Facebook;
- potranno essere visualizzate se qualcuno effettua una ricerca su Facebook o su un motore di ricerca pubblica;
- saranno accessibili per i giochi integrati in Facebook, le applicazioni e i siti Web utilizzati da te e dai tuoi amici;

A volte non potrai selezionare un pubblico quando pubblichi qualcosa (ad esempio quando scrivi sulla bacheca di una Pagina o commenti un articolo di giornale che usa il plug-in dei commenti). Ciò è dovuto al fatto che alcuni tipi di notizie sono sempre pubbliche. Come regola generale, considera che, se non vedi un'icona di condivisione, le informazioni saranno condivise pubblicamente.

Quando gli altri condividono informazioni su di te, possono anche scegliere di renderle pubbliche.

Informazioni sempre disponibili pubblicamente

Le informazioni elencate di seguito sono sempre pubbliche e, di conseguenza, trattate come informazioni che l'utente ha deciso rendere pubbliche.

- Nome
- Immagini del profilo e foto di copertina
- Rete
- Sesso
- Nome utente e ID utente

Utilizzo da parte di Facebook delle informazioni ricevute

Facebook utilizza le informazioni ricevute in relazione ai servizi e alle funzionalità offerte a te e ad altri utenti come i tuoi amici, i nostri partner, gli inserzionisti che acquistano inserzioni sul sito e gli sviluppatori che creano giochi, applicazioni e siti Web da te utilizzati. Ad esempio, Facebook può utilizzare le informazioni ricevute su di te:

- come parte dell'impegno costante a preservare la sicurezza di Facebook;
- per proteggere i diritti o la proprietà di Facebook o di altri;
- per mettere a tua disposizione funzioni e servizi sulla posizione, ad esempio facendo sapere a te e ai tuoi amici se c'è qualche evento in corso nelle vicinanze;
- per misurare o comprendere l'efficacia delle inserzioni visualizzate da te o da altri e mostrare inserzioni pertinenti;

- per fornire suggerimenti a te e agli altri utenti di Facebook, ad esempio: suggerire a un tuo amico di utilizzare lo strumento di importazione dei contatti che ti è servito a trovare degli amici, suggerire a un altro utente di aggiungerti agli amici perché tale utente ha importato il tuo stesso indirizzo e-mail, suggerire a un tuo amico di taggarti in una foto che ha caricato e in cui tu compari;
- per le operazioni interne, fra cui la risoluzione dei problemi, l'analisi dei dati, i test, la ricerca e il miglioramento del servizio.

Sebbene tu consenta l'utilizzo delle informazioni ricevute su di te da parte di Facebook, il proprietario di tali informazioni sarai sempre tu. La tua fiducia è molto importante, ecco perché Facebook non condivide con altri le informazioni ricevute su di te, se non nel caso in cui:

- abbia ricevuto la tua autorizzazione;
- ti abbia avvisato, ad esempio nelle normative del sito;
- abbia rimosso il tuo nome o qualsiasi altra informazione personale dal sito.

Memorizziamo i dati per il tempo necessario a fornire i prodotti e servizi a te e ad altri, compresi quelli descritti sopra. Tipicamente, le informazioni associate al tuo account saranno mantenute fino a quando il tuo account non sarà eliminato. Per alcune categorie di dati, ti informeremo sulle procedure specifiche di conservazione.

Un occhio critico a Facebook

Come mostrato sopra Facebook spiega per filo e per segno, addirittura con esempi, quale tipo di dato viene memorizzato, quali sono le caratteristiche dell'informazione condivisa e come possono essere usati i dati. Quindi riassumendo un individuo che conosce tutte queste regole può condividere informazioni su di se senza temere per la propria *privacy* o comunque rimanendone l'unico responsabile, ma cosa succede se sono gli altri a condividere delle informazioni sull'utente è tutto un altro punto.

Il regolamento dice chiaramente che le persone possono condividere informazioni su di te e decidere se renderle pubbliche o meno, il fatto preoccupante non è il primo aspetto, ma il secondo. Gli altri utenti possono avere informazioni sull'utente, ma non dovrebbe essere lui a decidere se renderle pubbliche o meno. Questo aspetto raggiunge una certa importanza se si ritorna al punto che adesso Facebook è indicizzato dai motori di ricerca e che quindi l'informazione fa parte del Web.

Un aspetto un po' "particolare" di questo regolamento è che in alcuni casi l'utente non potrà scegliere se un'informazione sia pubblica o meno, il sistema la metterà pubblica automaticamente.

C'è però anche un aspetto molto positivo in quello che Facebook spiega, l'utente può decidere di disiscriversi e di eliminare le sue informazioni personali dal database. Ovviamente chi ha scaricato le informazioni dell'utente in precedenza se le tiene.

Ciò che questo regolamento non dice è che se un'azienda vuole utilizzare i dati dell'utente, questa deve realizzare un'applicazione (es: *FarmVille*) poi l'utente deve autorizzare l'applicazione ad accedere ad alcuni dei propri dati personali, la pagina di accettazione mostra quali dati verranno presi dal profilo.

Twitter e la sua politica

«The future belongs to people who see possibilities before they become obvious.»

Ted Levitt

Queste sono alcune delle parole che s'incontrano nelle prime righe del sito Web della *Obvius Corporation* (Le frasi cambiano dinamicamente), l'azienda che ha creato Twitter. Ted Levitt era un professore ed economista americano ritenuto il fondatore del termine *Globalizzazione*.

« *We believe in the power of building systems that help people work together to make the world smarter, healthier, and just...better.* »

Obvious.com

Molte volte i giornali si parlano genericamente di *Social Network* come Facebook e Twitter senza rendersi conto che questi due sistemi hanno una architettura, uno scopo, un target di utenti e un tipo di dati assolutamente diversi tra di loro.

Facebook cerca di ricostruire il profilo di una persona con testo, immagini, video, amicizie, interessi e opinioni. Twitter invece permette al singolo utente di dire una cosa che sia visibile per tutti gli utenti della rete, o con chi decide lui, senza inserire troppe informazioni extra.

Una grandissima differenza strutturale tra i due sono le relazioni. Twitter usa un grafo diretto mentre Facebook no. L'utente A può seguire l'utente B anche se B non lo conosce e non sa niente di lui.

Su Twitter comunica con gli altri mentre su Facebook si mostra agli altri cosa siamo, o cosa vogliamo far credere di essere.

«L'utente sarà responsabile del proprio utilizzo dei Servizi, dei Contenuti postati sui Servizi e di ogni conseguenza derivante da tali azioni. I Contenuti inviati, postati o resi visibili dall'utente saranno visualizzabili da altri utenti dei Servizi e tramite servizi e siti web di terzi (accedere alla pagina "Impostazioni dell'account" per controllare chi può visualizzare i Contenuti). L'utente dovrà fornire esclusivamente Contenuti che ritiene di poter condividere con altri nel rispetto delle presenti Condizioni.»

Condizioni d'uso- Twitter

Twitter ha un approccio alla *privacy* molto differente da Facebook, il messaggio è «Quello che dici su Twitter potrà essere visto istantaneamente in tutto il mondo. Sei quello che twitti!»¹⁴.

Un'altra differenza è nel modo in cui le informazioni sono disponibili. Twitter usa delle *Api* per permettere l'accesso all'informazione, alcune richiedono un login altre no.

Attraverso queste è possibile raccogliere informazioni su tutta la rete senza bisogno di utilizzare un'applicazione come per Facebook. I limiti imposti da Twitter sono principalmente sulla quota oraria disponibile, quando si supera non è possibile fare altre richieste fino alla fine dell'ora.

Un occhio critico tra Facebook e Twitter

Pensando alla politica sulla *privacy* di Twitter si può assumere che il dato prodotto (il *tweet*) nasce espressamente per essere condiviso il più possibile e viene visto come una risorsa comune, visto che in genere qualsiasi utente può diventare nostro *follower* e averne accesso.

Il fatto che le relazioni non siano necessariamente corrisposte permette una maggiore flessibilità per la scelta del proprio pubblico e enfatizza le relazioni di Sorgente-Ripetitore che si generano con il sistema di *Retweeting*.

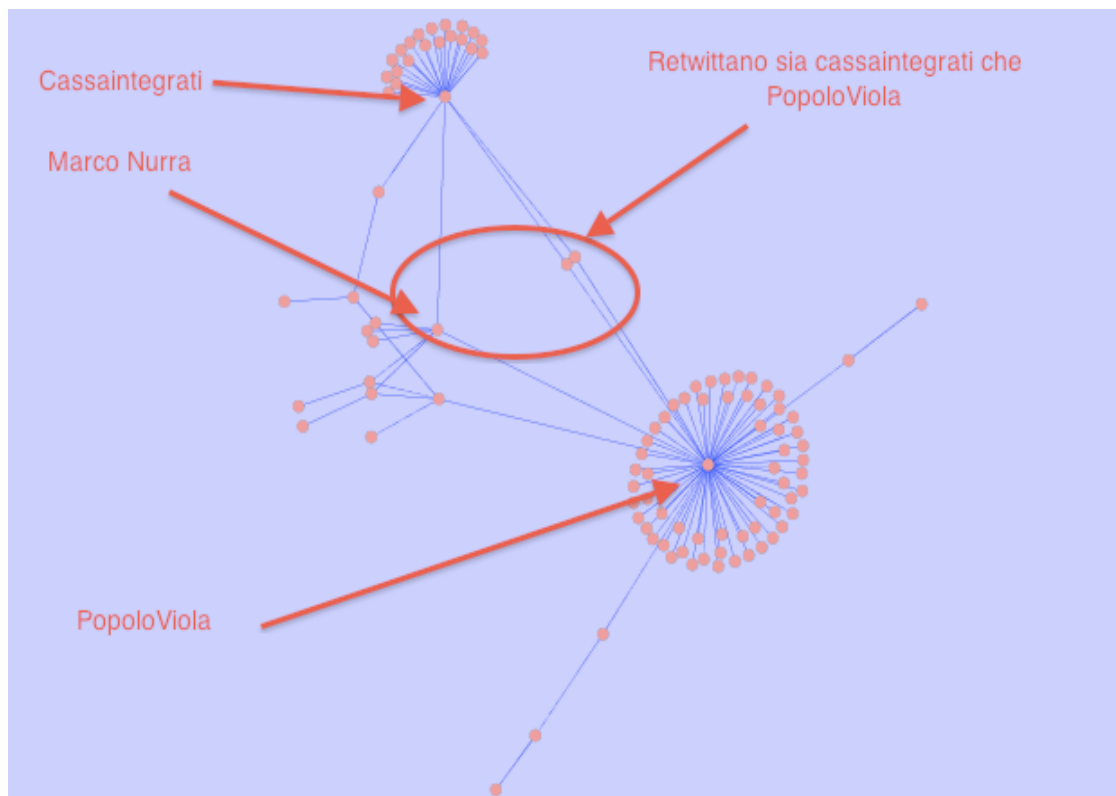
¹⁴ <https://twitter.com/tos>

14. Una storia sociale

Una delle principali differenze tra Twitter e Facebook è la percentuale d'informazione resa pubblica. Facebook per sua natura è un sistema per condividere informazione con gli amici. Twitter invece concede la possibilità alle persone di dire quello che pensano o sentano senza mettere un limite all'audience. Sono poi gli altri utenti a scegliere se valga la pena seguire o meno.

Twitter è un potentissimo mezzo di espressione personale grazie alle sue grandi dimensioni e alla sua struttura priva di limitazioni, chi sceglie di condividere qualcosa su Twitter si mette in gioco in un comunità dove tutti possono leggere e replicare. La comunità Twitter è allenata nella sintesi visto che i messaggi non devono superare i 140 caratteri, anche se molti ricorrono ai link esterni, e gli *hashtag* permettono di riferirsi con una sola parola un intero ambito di discussione. Questi *hashtag* sono una caratteristica molto importante della comunicazione su Twitter, come già visto nei capitoli precedenti. Il funzionamento di Twitter non è complicato, questo ha permesso di poter esprimere la propria opinione anche ai non esperti, coloro che avevano le cose da dire ma non possedevano gli strumenti per dirle.

Per molte aziende Twitter funziona come un servizio clienti dove è possibile porre domande e ricevere risposte. Le aziende hanno molto da guadagnare dall'uso di Twitter, sia come informazione diffusa che come informazione raccolta. In questo elaborato è stato mostrato il caso in cui durante il mese di Aprile un piccolo numero di utenti ha incominciato a parlare di come l'azienda stesse licenziando molti dipendenti anche se con un buon livello di anzianità. I due principali utenti che hanno diffuso idee su questo argomento sono stati **@cassintegrati** e **@popoloviola**.



Come si vede dal grafico **@popoloviola** ha molti più *retweet* di **@cassintegrati**, i messaggi di questi due utenti sono stati entrambi retwittati da alcuni utenti tra cui **Marco Nurra**. Questo utente è il fondatore del blog “l’isola dei cassintegrati¹⁵”. Guardando la rete Utenti-Hashtag del mese di Aprile, trattata nel capitolo 8, si nota che gli *hashtag* #ultimaparola, #sapevatelo hanno un grado tra i più alti e l’utente **Il_popolo_Viola** ha un valore di centralità abbastanza alto, è il 18° in classifica. Il nodo #ultimaparola ottiene anche un alto valore di *PageRank* posizionandosi quinto tra i venti nodi con *PageRank* più alto. Detto questo, anche il nodo #benetton ha dei valori di centralità, di grado e di *PageRank* molto alti, ma questo è un *hashtag* molto diffuso e molto probabilmente una parte rilevante delle sue occorrenze non è collegata con la critica sui licenziamenti. Comunque è giusto fare notare il peso dell’*hashtag* #benetton visto il messaggio di **@popoloviola** :

«76 lavoratrici buttate sulla strada, liquidate anche dopo 20 anni di servizio in Autogrill (dei #Benetton) che NON è in crisi. Gravissimo.»

E quello di **cassintegrati** :

¹⁵ <http://www.isoladeicassintegrati.com/>

«La famiglia #Benetton non solo non ritira i licenziamenti #Autogrill ma li raddoppia! 76 lavoratrici tra Roma, Milanoe Bologna.»

Il_popolo_Viola e **@popoloviola** fanno riferimento al solito utente; Twitter, come Facebook, cita un utente con un nome diverso dal *nick* in alcuni casi. Tramite alcune *Api* di Twitter è possibile sapere a chi appartiene un riferimento. Questa procedura è utilizzata principalmente per eliminare gli spazi nei nomi che risulterebbero fastidiosi nell'analisi e nella ricerca

Conclusione

L'obiettivo che è stato proposto nell'introduzione **“raccoliere ed elaborare dati sociali così che permettano l'individuazione di eventi sociali rilevanti e/o opinioni condivise”** ha richiesto la ricerca della sorgente di dati più adatta (Twitter), lo studio delle sue politiche sulla privacy e lo sviluppo di strumenti adatti. Grazie a questo processo di sviluppo è stato possibile archiviare ed elaborare informazione utile per rispondere alla domanda «Quale è la reputazione online di Autogrill?». Nel capitolo 14 viene riassunto un evento rilevante per l'azienda e la sua reputazione online.

Questa domanda è stato lo strumento per mettere alla prova le tecniche e gli strumenti sviluppati in questo lavoro di ricerca iniziato nel Gennaio del 2012.

Sono stati estratti gli utenti più rilevanti per l'argomento, gli *hashtag* più vicini all'azienda e diversi eventi sociali critici per il marchio (es: la protesta contro i licenziamenti). Tutto questo è stato fatto nel periodo di sviluppo di strumenti e non attraverso l'uso di strumenti consolidati disponibili sul mercato. Il grafo ottenuto con questi dati ha portato a una maggiore comprensione della rete di idee che si sono formate intorno al marchio Autogrill e ai suoi prodotti e, l'analisi di questo grafico, può essere un importante strumento per basare una strategia di *Marketing* e/o di comunicazione. I dati raccolti hanno il vantaggio di essere rilasciati in maniera spontanea e non sotto forzatura, come può succedere in un sondaggio dove le persone vengono intervistate.

La raccolta delle informazioni da Twitter e non da Facebook è stata decisa dopo l'analisi delle due diverse politiche di *privacy* le quali incarnano a pieno lo spirito assai differente dei due servizi. Twitter parte dal principio che ciò che condividi lo condividi con il mondo, Facebook invece assume che si possa condividere delle informazioni solo con le persone da noi decise, più la società Facebook stessa. La differenza non sta solo nella diffusione delle informazioni dirette, ma anche di quelle non dirette. Su Facebook i nostri amici possono condividere informazioni su di noi decidendo da soli se siano pubbliche o meno, quindi il problema non è

più solo come ci comportiamo noi ma anche come si comportano gli altri. Possiamo decidere di eliminare l'informazione inserita da altri solo quando questa è stata inserita, in questo caso Facebook ha già archiviato il dato e sarà difficile cancellarlo perché è un'informazione che non appartiene a noi, ma a chi la condivide.

Come già detto nell'introduzione questo lavoro non vuole esaurire la conoscenza del settore anzi, un obiettivo implicito è quello di portare parte delle informazioni raccolte, delle esperienze, ma soprattutto delle opinioni per **stimolare il dibattito su cosa si possa fare con il dato sociale e su cosa non si possa fare**; poi in un secondo momento, quando si è decisa una politica di riservatezza adeguata, si può passare al **come farlo**.

Decidere come fare questo tipo di lavoro è veramente complicato e dipende principalmente da cosa si cerca e cosa ci si vuole fare, un esempio di questo è il fatto che i dati estratti per questo elaborato siano basati sull'utilizzo di parole chiave e che questa sia solo una delle tre strategie discusse in questo elaborato.

Glossario

Su Twitter

- **Tweet:** messaggio scritto su Twitter. Può contenere al massimo 140 caratteri. Se raggiunge la lunghezza massima viene detto *Twoosh*;
- **Retweet:** Quando un utente condivide un *tweet*;
- **Via:** quando un utente condivide un *tweet* utilizzando un'interfaccia differente da quella del suo profilo. Per esempio è possibile condividere *tweet* da *Youtube* o dai siti di alcuni giornali.
- **Hashtag:** termine preceduto dal simbolo "#", su Twitter assume un ruolo centrale nella comunicazione.

Analisi delle reti

- **Nodo:** si riferisce al singolo attore rappresentato nella rete;
- **Arco:** Esprime una relazione tra due nodi;
- **Grado del nodo:** indica quanti archi sono collegati al nodo;
- **Centralità di un nodo:** indica quanto il nodo è centrale nella rete;
- **Pagerank:** è un valore che prova a stimare l'autorevolezza di un nodo basandosi su chi è collegato ad esso;

Bibliografia

David Easley, Jon Kleinberg: *Networks, Crowds, and Markets*. Cambridge University Press (2010);

Russell M. A. : *Data mining nel social web*, O'Reilly (2011);

Heerschop, B., et al: *Polarity Analysis of Texts using Discourse Structure*, 2011;

Esuli, A., Sebastiani, F., *SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining*;

Pigni F., Ravarini A., Sciutto D. : *Sistemi per la gestione dell'informazione*, Apogeo (2009);

Gonzalez-Ibanez, R., et al: *Identifying Sarcasm in Twitter: A Closer Look*, Proceedings of the 49° Annual Meeting of the Association for Computational Linguistics: shortpapers, pagg 581-586, 2011;

Xie, X.: *Potential Friend Recommendation in Online Social Network*, 2010 IEEE/ACM International Conference on Green Computing and Communications & 2010 IEEE/ACM International Conference on Cyber, Physical and Social Computing, 2010;

Jiang, L., et al: *Target-dependent Twitter Sentiment Classification*, Proceedings of the 49° Annual Meeting of the Association for Computational Linguistics: shortpapers, pagg 151-160, 2011;

Baccianella, S., et al: *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*;