

ROMA
21 Novembre 2013

BIG DATA E ANALISI PREDITTIVA NEL MONDO BANCARIO

Institutions
Markets
Technologies

IMT

INSTITUTE
FOR ADVANCED
STUDIES
LUCCA

**IMT
ALTI STUDI LUCCA**

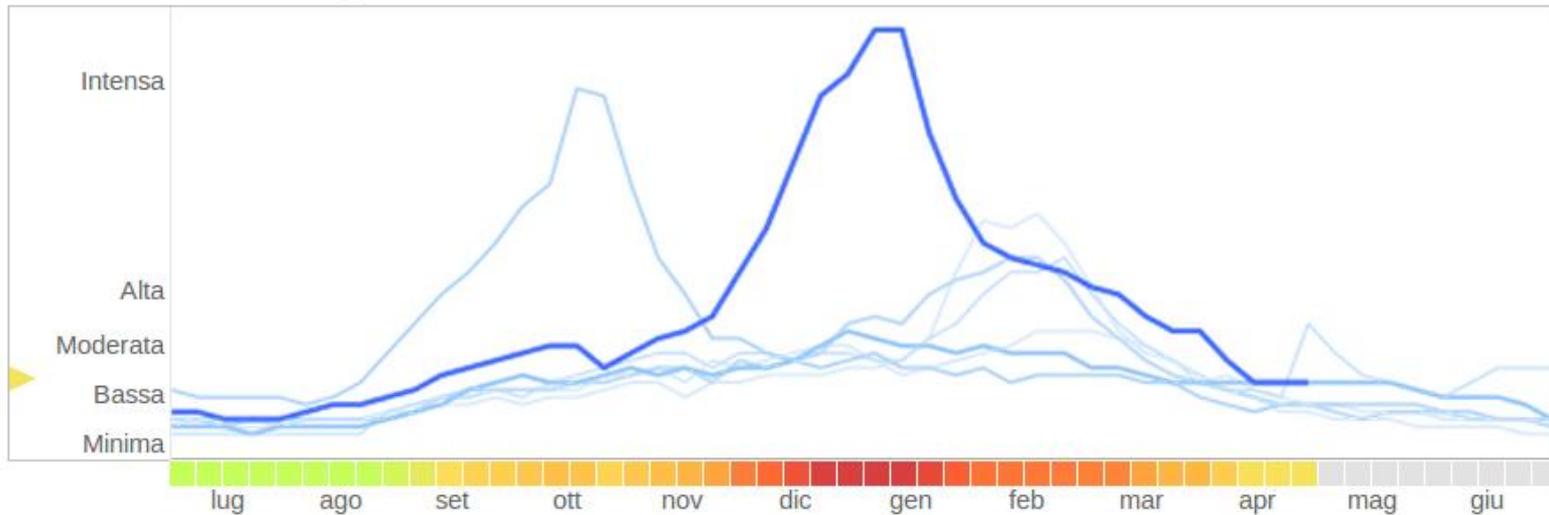
MASSIMO RICCABONI

Big Data 'Power':

Stati Uniti > Virginia

google.org Trend influenzali

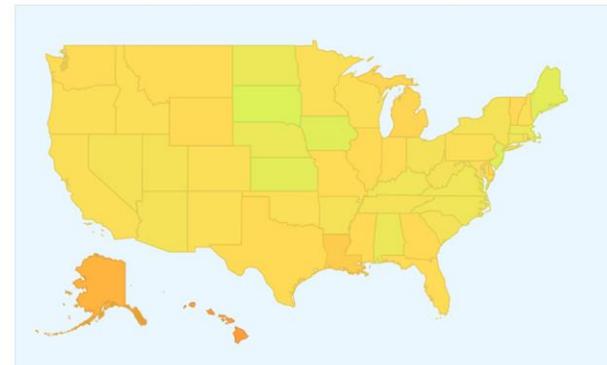
● 2012-2013 ● Anni passati ▼



2007-2008 U.S. Flu Activity - Mid-Atlantic Region

ILI percentage

● Google Flu Trends ● CDC Data



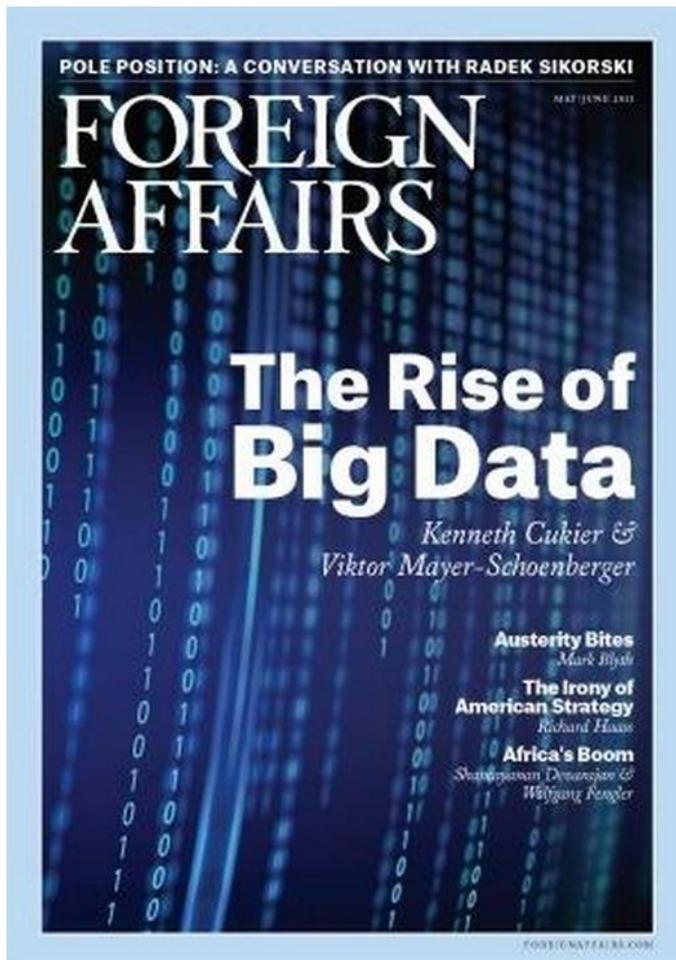
ROMA
21 NOVEMBRE 2013

MASSIMO RICCABONI

BIG DATA E ANALISI PREDITTIVA
NEL MONDO BANCARIO

Institutions
Markets
Technologies
IMT
INSTITUTE
FOR ADVANCED
STUDIES
LUCCA

Big Data nelle news:



Google Trends per 'Big Data':



Trends

Web Search Interest: **big data**. Worldwide, 2004 - present.



Explore trends

Hot searches

Search terms

big data

+ Add term

▶ Other comparisons

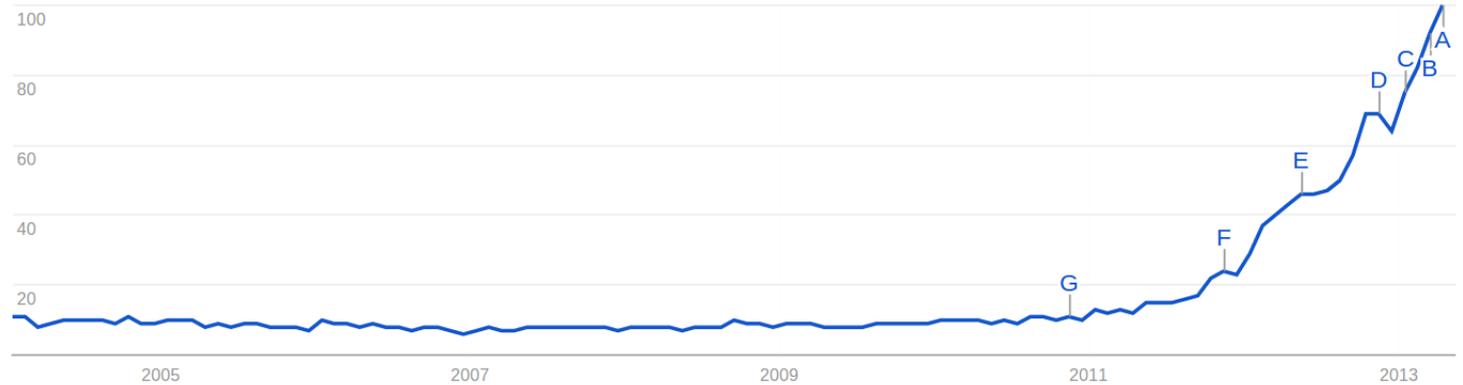
Limit to

Web Search

Interest over time

The number 100 represents the peak search interest

News headlines Forecast



Big Data in Social Media

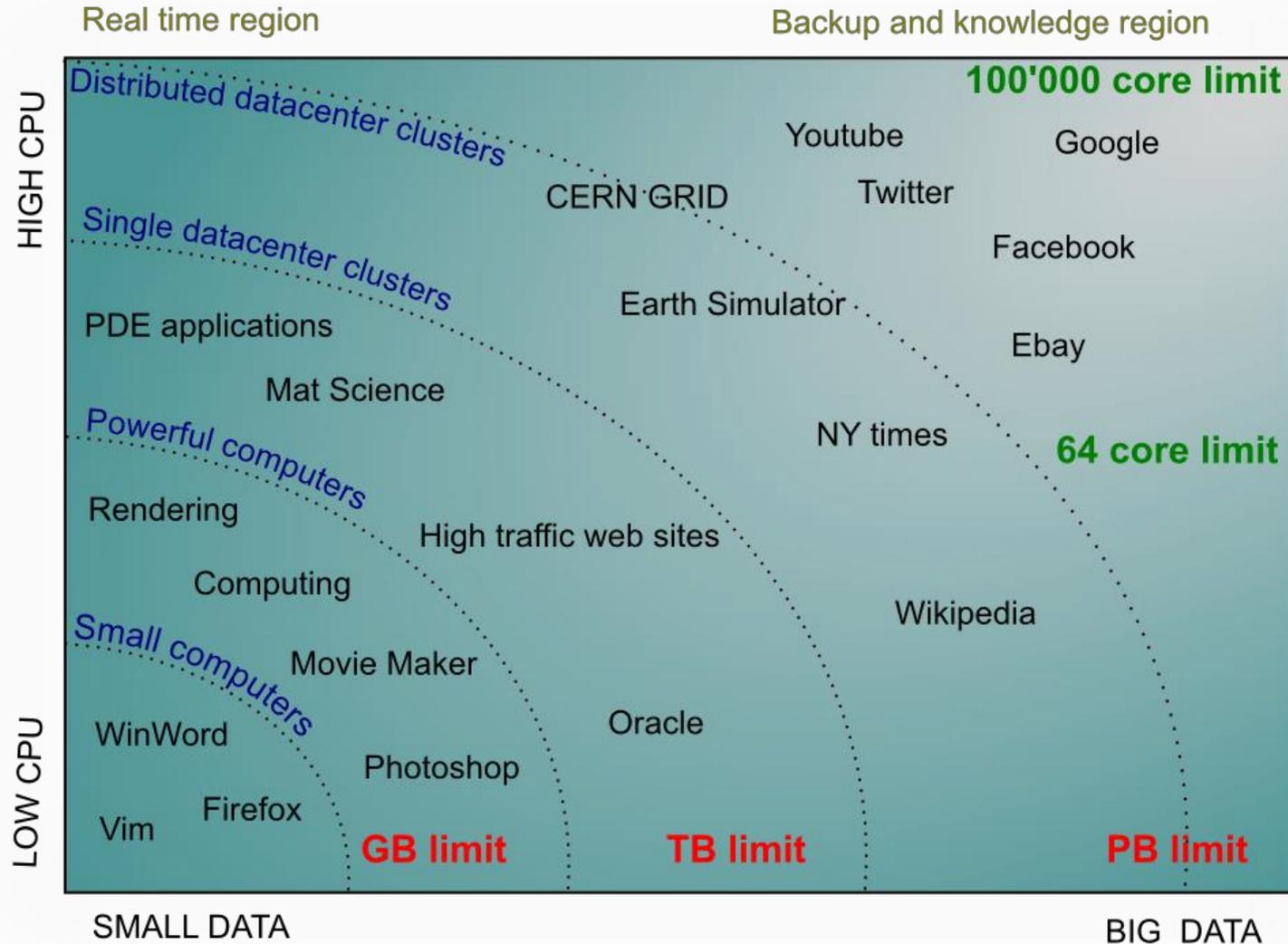
Facebook is collecting your data — 500 terabytes a day

by [Eliza Kern](#) AUG. 22, 2012 - 3:25 PM PDT

 10 Comments    +1 

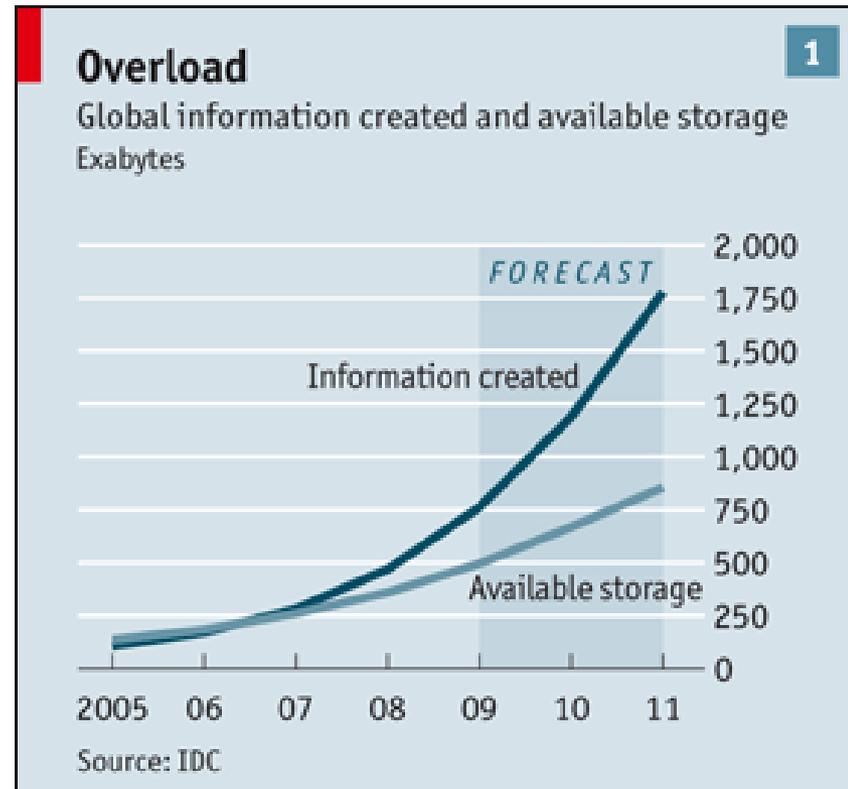
- 2.5 billion content items shared per day (status updates + wall posts + photos + videos + comments)
- 2.7 billion Likes per day
- 300 million photos uploaded per day
- 100+ petabytes of disk space in one of [FB's largest Hadoop](#) (HDFS) clusters
- 105 terabytes of data scanned via Hive, Facebook's Hadoop query language, every 30 minutes
- 70,000 queries executed on these databases per day
- 500+terabytes of new data ingested into the databases every day

La galassia dei Big Data:



Problema dell'archiviazione dati:

L'informazione creata cresce molto più velocemente dello spazio di archiviazione disponibile



Alla scala dei **Terabyte**...:

RAM memory < 32GB ... data on hard disk 1TB and more

Inside the cpu ...only 6MB of high speed memory

Network ? 100Mb/sec

Hard disk ? 200Mb/sec I/O

No relational joins are possible

The end of RDBMS and SQL ? (the No-sql revolution)

To Sort or Not to Sort ? (the Key / Value paradigm)

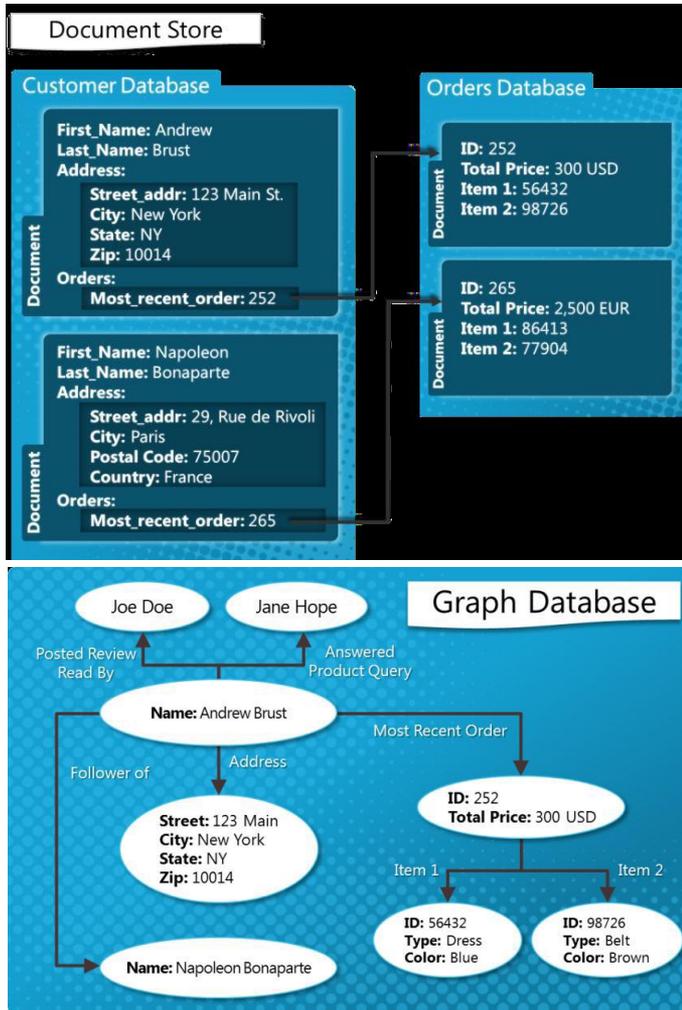
Big Data Tools :

Nuovi paradigmi di calcolo e di storage

- **Map/Reduce** paradigm : originally from Google
- **Divide et impera** ... big problem in many small problems in parallel
- **Key/value** technique in map reduce operations
- **Hadoop** (framework per la computazione)



La rivoluzione no-SQL:



Il diluvio dei dati e la fine della scienza:

"All models are wrong, but some are useful." (George Box 30 years ago)

"Consider physics: newtonian models were crude approximations of the truth (wrong at the atomic level, but still useful)...

...statistically based quantum mechanics offered a better picture — but quantum mechanics is yet another model, and as such it, too, is flawed, no doubt a caricature of a more complex underlying reality...

...then physics has drifted into theoretical speculation about n-dimensional grand unified models over the past few decades (the "beautiful story" phase of a discipline starved of data) is that we don't know how to run the experiments that would falsify the



WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

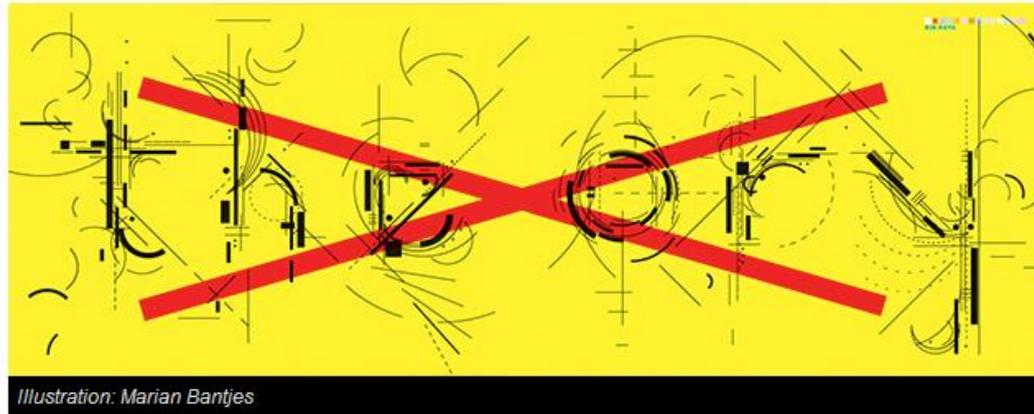


Illustration: Marian Bantjes

"But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete."

"...companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all."

"...Google conquered the advertising world with nothing more than applied mathematics. It didn't pretend to know anything about the culture and conventions of advertising — it just assumed that better data, with better analytical tools, would win the day. And Google was right."

"Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any

La nuova Scienza delle Reti :



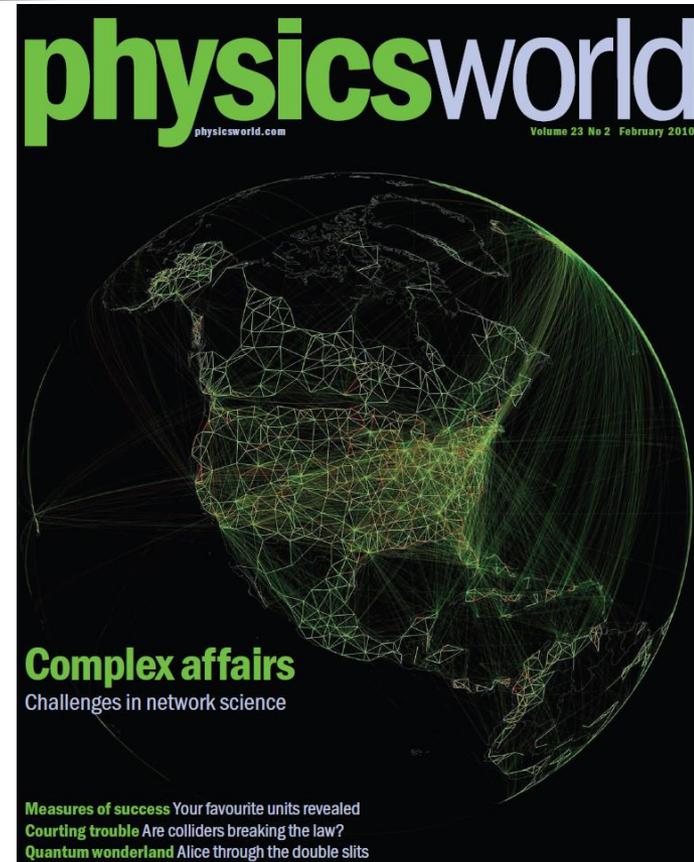
"Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves."

"Correlation is enough. We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."

Our view:

In a way through a 'theory' you try to guess data you don't have, starting with a data sample and a model...

...but if you have a huge amount of data, you have almost all the trajectories and you have only the squeeze information from the data you have, choosing the right paths. But there is a new science, data mining oriented, to leverage from... (Network Science)



Complexity: Network science

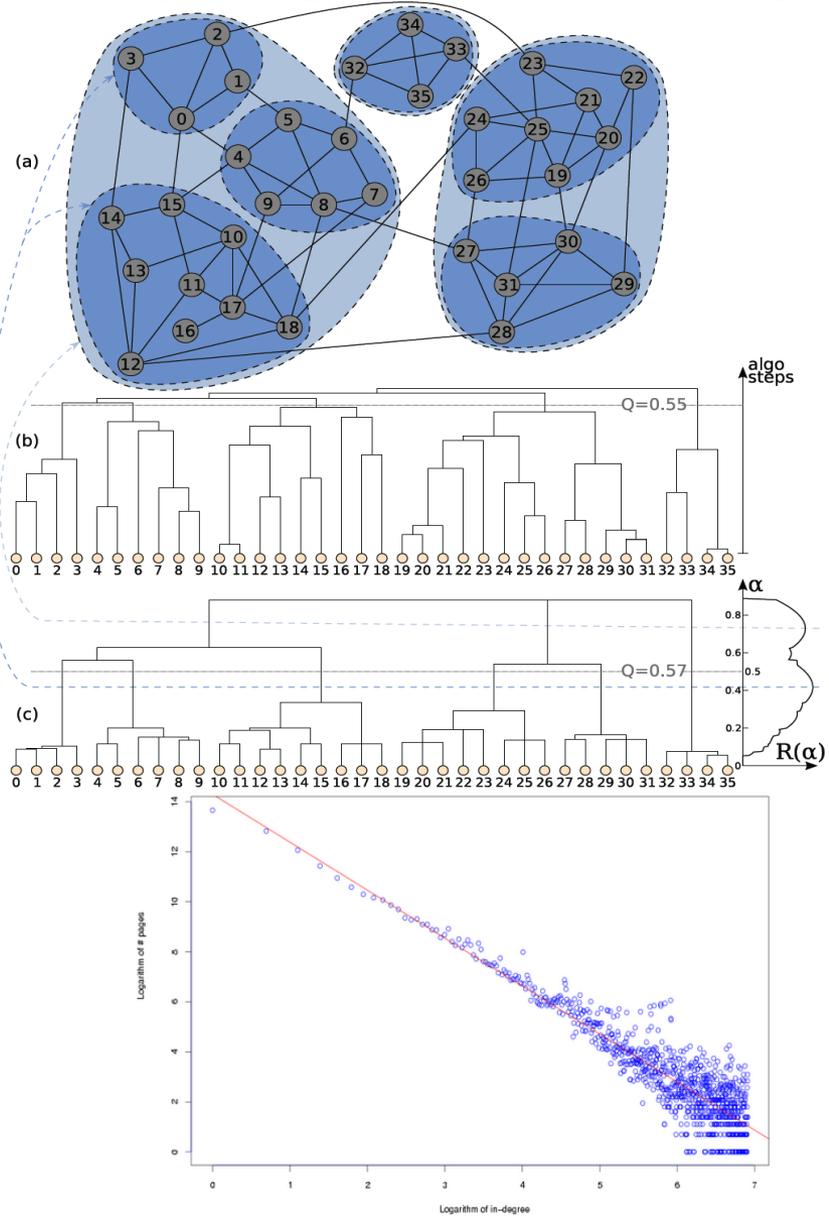
A networked world

Mark Buchanan and Guido Caldarelli chart the remarkable rise of network science, examine why it has become so popular, and predict what its future may bring

Big Data Scenario

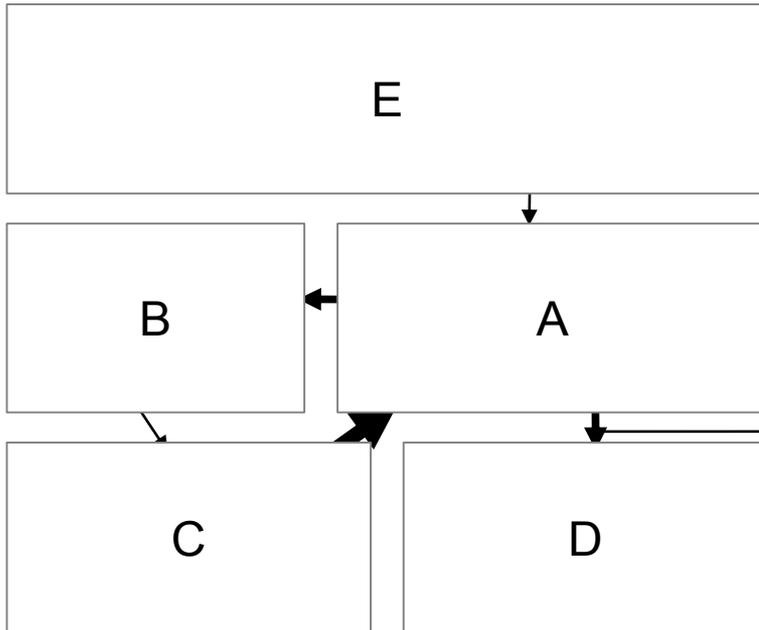


PB
data

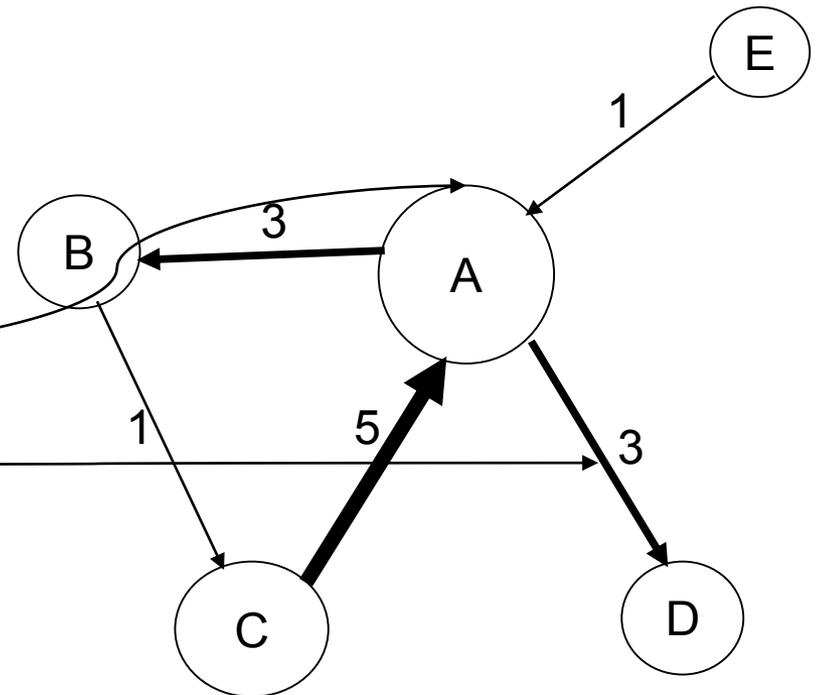


Il valore aggiunto delle *Complex Networks*:

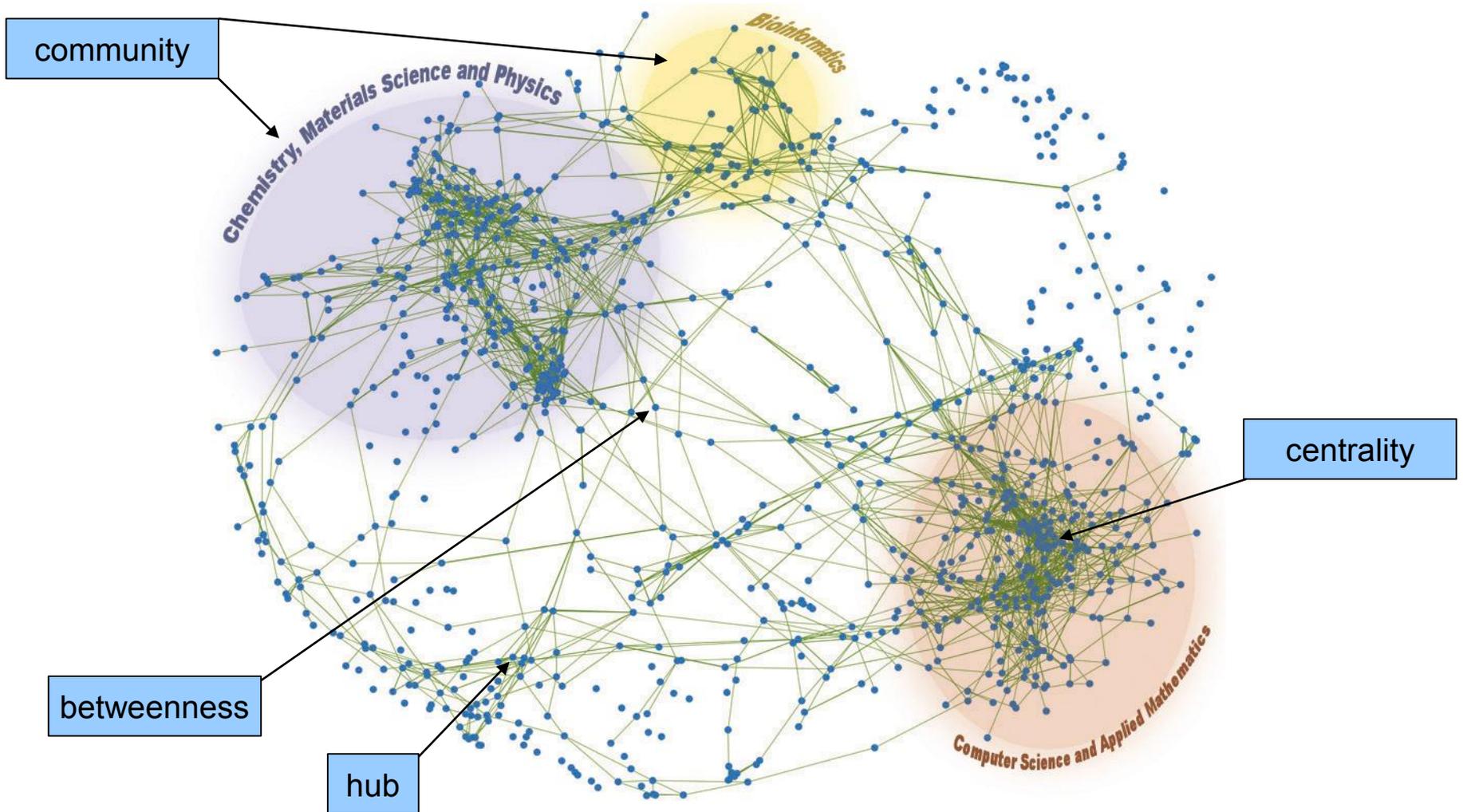
SISTEMA COMPLESSO



NETWORK

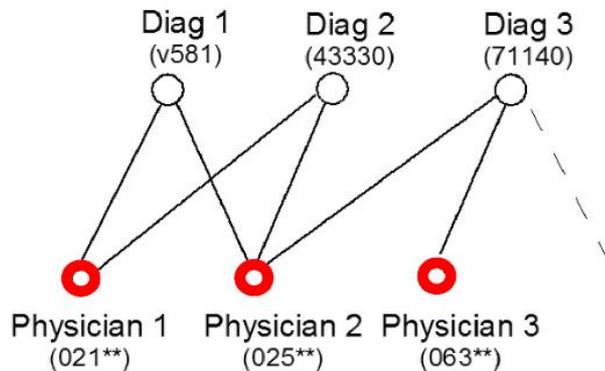


Estrarre informazioni dalle Reti:

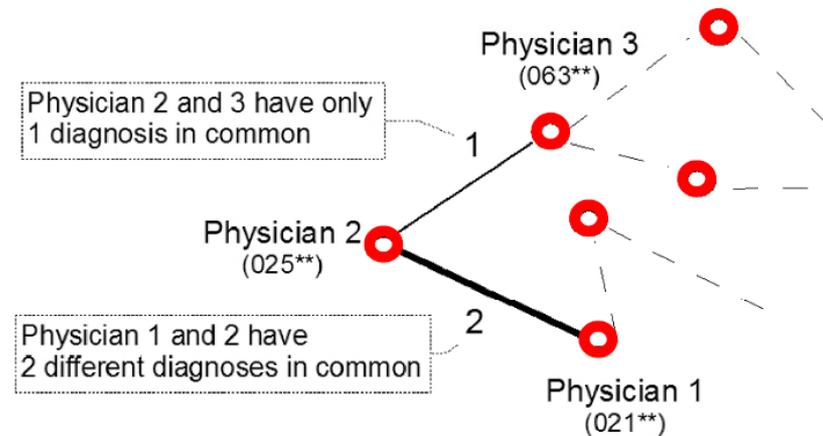


Reti Complesse da Database:

PATIENT	GENDER	ADMISS.	DISM.	DIAG.	PROT.	DRG	PHYS/SURG	COST	DEPART.
111	1	06/09/03	06/12/03	07054	8952	206	034**	€ 1.728	Dep. 2
112	1	09/03/03	09/15/03	71140	8838	242	025**	€ 2.059	Dep. C (Surgery)
110	2	03/04/03	03/11/03	9999	8874	468	034**	€ 4.660	Dep. 2
113	2	01/10/03	01/12/03	43330	8744	15	021**	€ 1.872	Dep. 2
113	2	01/30/03	02/01/03	V581	9925	410	021**	€ 1.411	Dep. 2
244	1	09/21/03	09/24/03	V581	9925	410	025**	€ 1.411	Dep. C (Surgery)
718	2	02/13/03	02/20/03	1970	9925	82	028**	€ 3.050	Dep. 2
715	1	06/30/03	07/05/03	56211	4513	183	013**	€ 1.397	Med. Lab. (MDL)
717	2	07/09/03	07/09/03	71140	8838	242	063**	€ 2.059	Dep. C (Surgery)
719	1	08/03/03	08/08/03	43330	3812	5	025**	€ 2.989	Dep. C (Surgery)



(a) Bipartite Network



(b) Unipartite Network

Dalla Fisica al WWW:

La Fisica Teorica del WWW

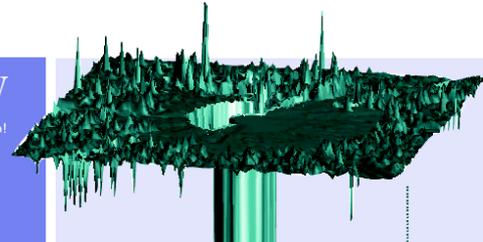
Linkalab, in collaborazione con il CNR-INFM di Roma e Yahoo! Research Lab, ha individuato una sorprendente connessione tra la Fisica Teorica e il PageRank, l'algoritmo di Google che misura la popolarità dei siti web. Secondo questa ricerca le pagine web del WWW possono essere viste come gli snodi di un sistema fisico descritto da un potenziale: esattamente come accade per un sistema quantistico o gravitazionale



↑ ↓
Yahoo! è un esempio di 'hub' del web con alto PageRank essendo citato (puntato) da molti siti importanti e meno importanti. In corrispondenza del minimo della buca di potenziale troviamo il massimo del PageRank.

PageRank

Il PageRank è un algoritmo che misura la popolarità di un sito in relazione alla popolarità dei siti vicini che lo puntano tramite un hyperlink. L'analogia consiste nell'aver trovato una forma di potenziale del web nella quale compaiono esplicitamente i numero di link entranti e uscenti in ciascuna pagina.

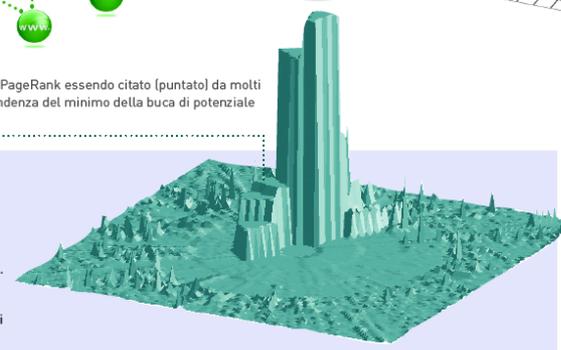
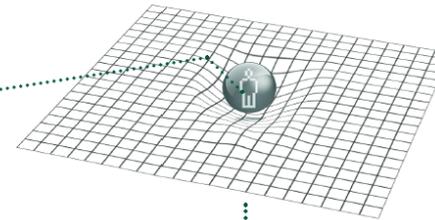


Buca di potenziale

Molti sistemi fisici possono essere caratterizzati da un profilo di potenziale. La dinamica di questi sistemi può essere descritta come una traiettoria di una pallina che viene deviata o cade nelle buche di questo potenziale.



↑ ↓
La pallina nel nostro caso rappresenta un utente Internet che esplora lo spazio web. Quando si avvicina a un sito importante ne è attratto e molto spesso 'cade' nella sua Home Page, proprio come avverrebbe in un sistema fisico.



La Fisica accelera anche il WWW

Questa analogia nelle equazioni apre dei nuovi scenari. Potrebbe essere possibile impiegare direttamente i risultati già noti nel campo della Meccanica Quantistica per predire l'evoluzione dei link (cioè per scoprire chi si conatterà con chi) e per quantificare l'attrattività "fisica" dei siti della Rete. Durante la ricerca è stata già introdotta una "scorciatoia" per il calcolo del PageRank che dimostra l'efficacia della scoperta. Il team di ricerca di Linkalab è in piena attività per esplorare i nuovi territori che si potranno raggiungere grazie a questo ponte tra Internet e la Fisica Teorica.

linkalab

ROMA
21 NOVEMBRE 2013

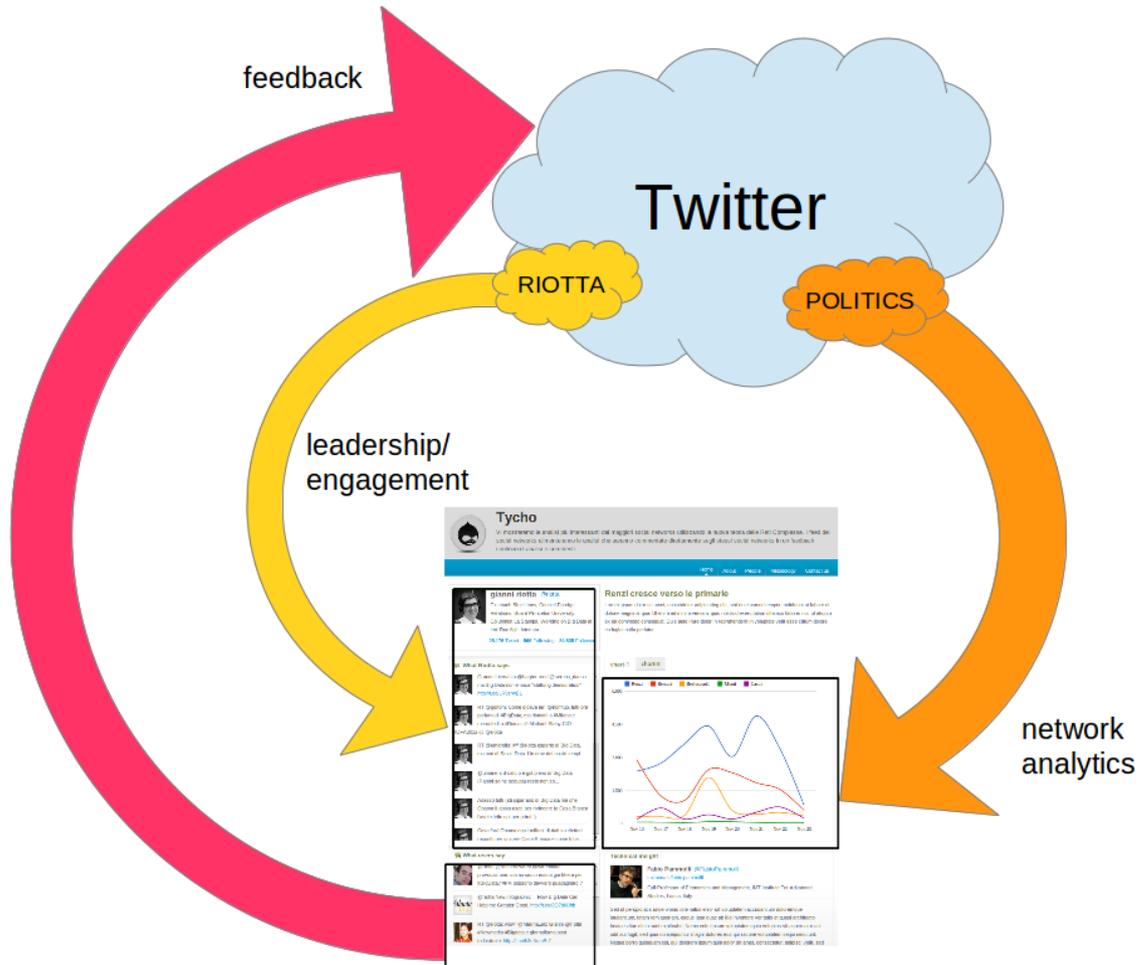
MASSIMO RICCABONI

BIG DATA E ANALISI PREDITTIVA
NEL MONDO BANCARIO

Institutions
Markets
Technologies
IMT
INSTITUTE
FOR ADVANCED
STUDIES
LUCCA

PROGETTO TYCHO BIG DATA

Tycho Streams:



Tycho Interface :

Home | Archivio | About | Contact us

Tycho | IMT *

“ Vi mostreremo le analisi più interessanti dai maggiori Social Network, utilizzando la nuova Teoria delle Reti Complesse. I flussi di 'tweet' e 'blog post' alimenteranno le analisi che saranno commentate sugli stessi Social Network, in un feedback continuo di analisi e commenti.

* Tycho è una piattaforma tecnologica di Social Analytics che nasce dalla collaborazione tra IMT Lucca e Linkalab.

Analista
Gianni Riotta @riotta

Ex coach Blue Lions, Council Foreign Relations, Board Princeton University, Columnist La Stampa. Working on Big Data at Imt. Due figli. Interista

32.237 Tweet 634 Following 108.822 Follower

Cosa dice Riotta

- RT @GuidoCaldarelli: @FabioPammolli @tychobigdata @riotta @riccabonim @cinziacaporale meglio meglio della casalinga :)
- RT @GuidoCaldarelli: @riotta @tychobigdata adesso che cominciamo a produrre grafi da Twitter vedrete quanto ci divertiamo
- RT @FabioPammolli: @GuidoCaldarelli @tychobigdata @riotta acclarata centralità maestro Vigevano e casalinga Voghera @RiccaboniM @cinziacaporale
- RT @Panna975: @riotta @tychobigdata sarà visione semplicistica ma chi usa TW con cognizione funziona, chi ci sta x dovere affonda #Democraziainrete
- @mattiaferraresi @tychobigdata @GuidoCaldarelli @mumon la parola (140 caratteri) ai fisici statistici
- RT @GuidoCaldarelli: @tychobigdata @riotta c'è ancora molto da scoprire nel grafo di Twitter. La rete

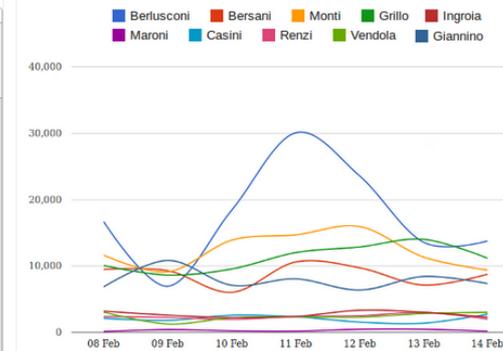
Cosa dice la rete

- RT @nomfup: Ma la Rete nujaaffatta (featuring @simonespetia @salernosal @giulianopisapia @3O0spartani @riotta @tychobigdata): <http://t.co/1qTGdZ3>
- @tychobigdata Elezioni: Bersani, Monti e Berlusconi si sfidano su Twitter <http://t.co/edMiSonF>
- Ma la Rete nujaaffatta (featuring @simonespetia @salernosal @giulianopisapia @3O0spartani @riotta

Il poker Bersani Berlusconi Grillo Monti a 7 giorni dal voto

A sette giorni dal voto Bersani rafforza la presenza politica e il vantaggio sul web ma Berlusconi dimostra capacità di protagonismo e rimonta. Malgrado il no all'intervista su Sky Beppe Grillo conferma la forza del suo appeal online mentre Monti ha le sue teste di ponte a Roma e Milano centro del dibattito politico economico. Città per città e hashtag per hashtag tycho mappa la conversazione social delle elezioni 2013 e scopre a sorpresa che anche la tv diventa digitale e interattiva.

Volumi Rete Mappa Berlusconi Mappa Monti Mappa Bersani Mappa Grillo



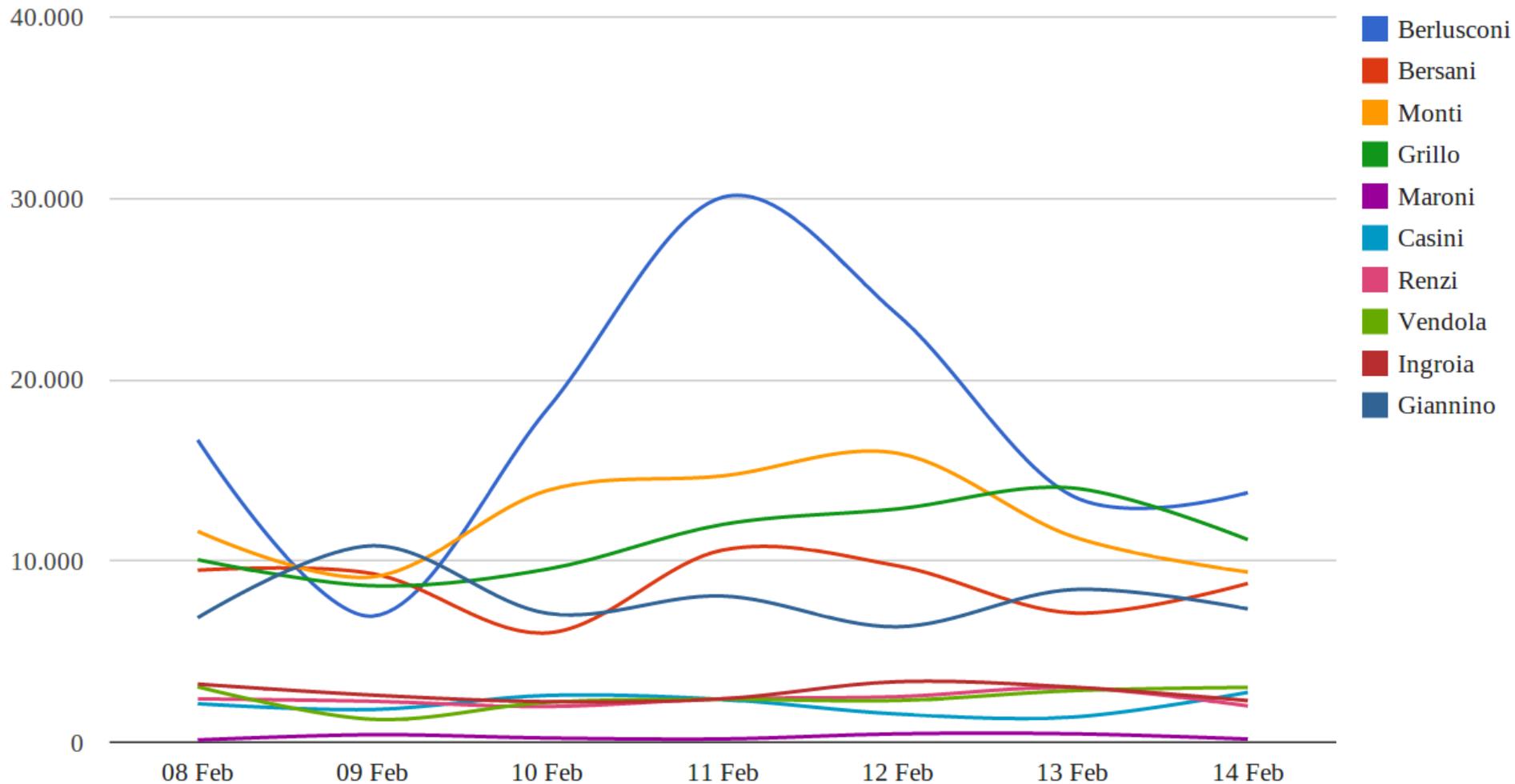
> Vai al grafico dei volumi dei politici

Punto Tecnico

Guido Caldarelli @GuidoCaldarelli
imtlucca.it/guido.caldarelli
Associate Professor in Theoretical Physics, IMT Institute For Advanced Studies, Lucca, Italy

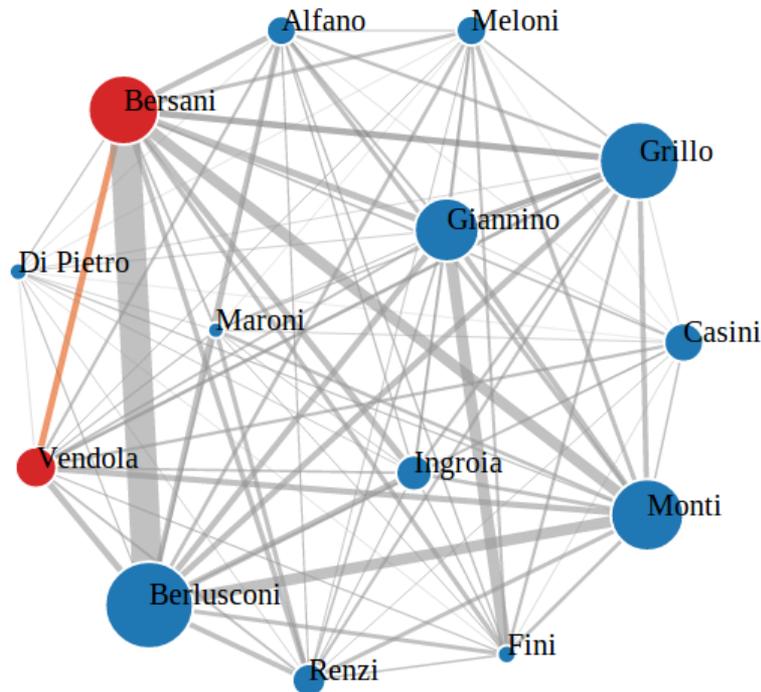
I dati di oggi presentano il volume di twitter per alcuni dei principali politici in corsa per le prossime elezioni. Ricordiamo che un tweet (cinguetto) è un messaggio di 140 caratteri in cui si possono indicare degli argomenti con un termine preceduto da un hashtag (#). Nel nostro caso abbiamo considerato tutti i messaggi riferibili sia ai candidati premier alle prossime elezioni, sia riferibili ad altre personalità politiche (in caso penerino un volume consistente di dati).

Tycho Charts:



Tycho Networks :

La rete dei politici italiani del 14 Febbraio 2013



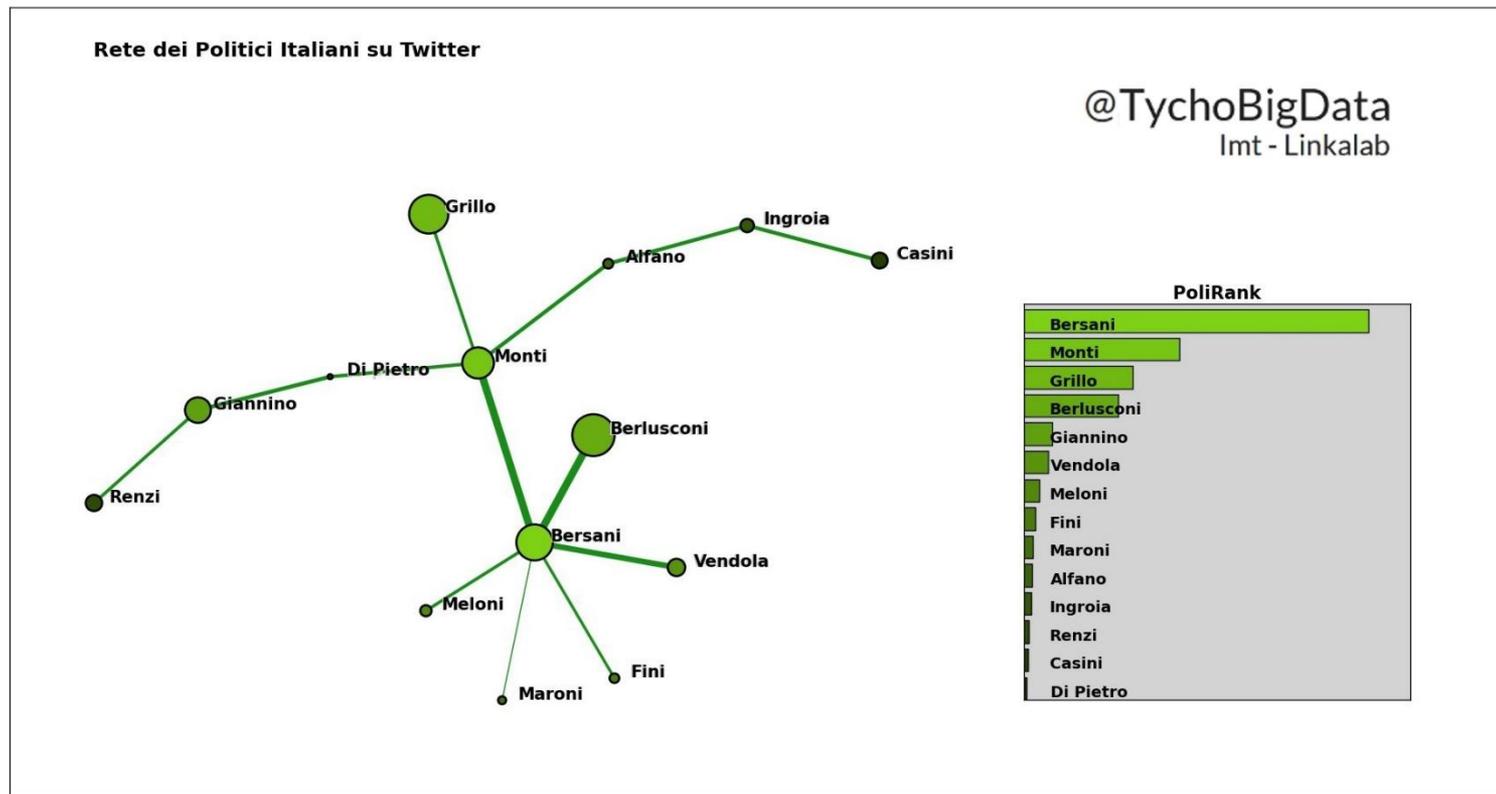
Vendola

ottoemezzo
elezioni2013
taranto
Calenda
Fitto
Ber
TgLa7
Sel
Ottoemezzo
tsunamitour
Elezioni
bersagliomobile
24Mattino
OttoEMezzo
SanValentino
Rai
Gruber
Bra
Cupido
stagflazione

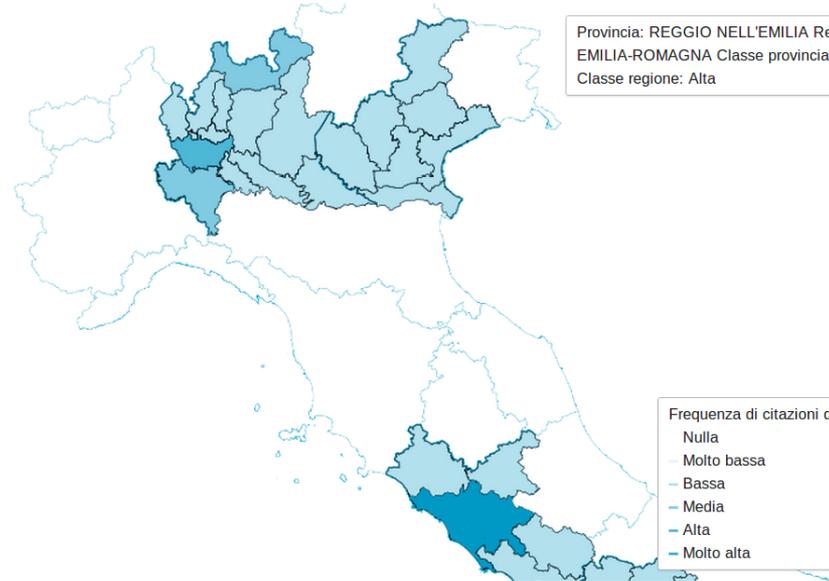
Bersani

agorarai
TgLa7
elezioni2013
Finmeccanica
M5S
lazzanara
taranto
Cagliari
Agca
Elezioni2013
Schauble
Grecia
ottoemezzo
Ber
Pistorius
BandaLarga
Tangenti
PressioneFiscale
bersagliomobile
onebillionraising

Tycho Ranking:



Tycho Maps :



Tycho La Stampa (10-02-2013):

LA STAMPA
QUOTIDIANO FONDATAO NEL 1867
DOMENICA 10 FEBBRAIO 2013 - ANNO LXXIV N. 40 - 1,20€ (IN ITALIA PREZZO PREDICAZIONE ED EDIZIONE IN ULTIMA SPERANZA ASS. PIZZALI - D.L. 33003 (2004) ART. 1 COMMA 1, D.C. - WWW.LASTAMPA.IT

Analisi Tycho-La Stampa sul dibattito in rete
Così le promesse dei leader accendono l'Italia di Twitter
Berlusconi decolla con l'Imu

Flusso di Twitter
In migliaia
Berlusconi
Mancini
Bersani
Cassini
Monti
Rosati

LA CAMPIAGNA FREDDA
IL VOTO D'INVERNO SVUOTA LE PIAZZE E RIEMPIE LE TV
MATTIA PIZZINI

Un Paese che non sa punire i razzismi
FRANCESCO BORGARE

LE IDEE
Un Paese che non sa punire i razzismi

Bankitalia chiede più poteri
Visco sul caso Montepaschi: "Una norma per rimuovere i cattivi manager"
Bersani attacca il premier: "Un demagogo e sul bilancio Ue vittoria di Pirro"

SOCIALISTI EUROPEI
"Elezioni italiane svoltate per l'Europa"
LUCA BORGARE

NEL MIRINO I 90 COMPAGNE
Energia, l'accusa dell'Authority "Robin tax scaricata sui clienti"
LUIGI GRASSI AINGUZZA

LA STORIA
Usa, per i reduci la guerra non finisce mai
PAOLO MANTOVANI

DOPO LA TEMPESTA IN AMERICA, L'ALLARME MALTEMPO IN ITALIA DOMANI E MARTEDI FERMA GLI ASPIRANTI DOCENTI
Arriva la neve, salta il concorso della scuola

LA LINEA DURA DEL GOVERNATORE DINTORNI BANCHEGGI
FRANCESCO MANZOCCHI

grandi banchieri che sono entrati ieri mattina in piazza d'Armi a un appuntamento con il governatore della Banca d'Italia non sarà di risuldo. Il banchiere Ignazio Visco sostiene che il caso Mps è isolato e che il sistema è sano.

LA STORIA
Usa, per i reduci la guerra non finisce mai

DOSSETTI
Il cristiano "morsicato" dal Vangelo
EUGENIO BIANCHI AINGUZZA

Colifagina
Regolarizza
La linea
Dopo il pasto
Dopo la notte

QUATTROVOTE
La prova della nuova Opel Adam

Acqua Eva, la sorgente più alta d'Europa, nasce dal Monviso. Prevalto: è tra le acque con meno sodio al mondo.

ABBONAMENTI | ARCHIVIO | PIU' VISTI | SOCIAL | METEO | TUTTOAFFARI | LAVORO | LEGALI | NECROLOGIE | SERVIZI |

LA STAMPA.it

Cerca...

TORINO - CUNEO - AOSTA - ASTI - NOVARA - VCO - VERCELLI - BIELLA - ALESSANDRIA - SAVONA - IMPERIA e SANREMO
VOCI DI MILANO

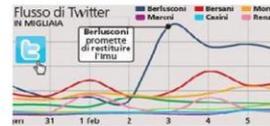
ATTUALITÀ | OPINIONI | ECONOMIA | SPORT | TORINO | CULTURA | SPETTACOLI | MOTORI | DONNA | CUCINA | SALUTE | VIAGGI | EXTR@ | FOTO | VIDEO

HOME | POLITICA | ESTERI | CRONACHE | COSTUME | TECNOLOGIA | SCIENZA | AMBIENTE | LAZAMPA | I TUOI DIRITTI | MARE | MONTAGNA | SPECIALI

la Feltrinelli.it Tutti i Libri subito disponibili al 15% di sconto INIZIA LO SHOPPING

Così le promesse dei leader accendono l'Italia "Twitter ci svela tutti i segreti delle elezioni"

Flussi, grafici e mappe/ Lo speciale sul voto



MARCO BARDAZZI, GIANNI RIOTTA
L'analisi Tycho-Lastampa sul dibattito in rete Berlusconi decolla dopo la proposta sull'Imu.
Le nostre elaborazioni a due settimane dal voto. Il fisico Chessa: «Fondamentale analizzare le connessioni. Incrociando tutte le informazioni potremmo scoprire anche chi ama il cioccolato»

SPECIALE Le voci della rete / Trova il tuo candidato

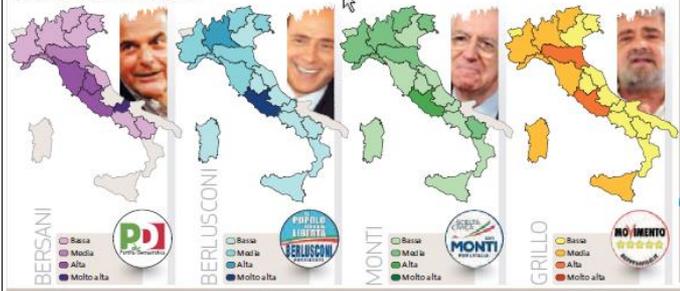


Tycho La Stampa (10-02-2013):

4 | Primo Piano | L'ESPRESSO | 21 NOVEMBRE 2013

-14 giorni al voto

Frequenza di citazioni per Regioni



Così le promesse dei politici accendono il dibattito in Rete

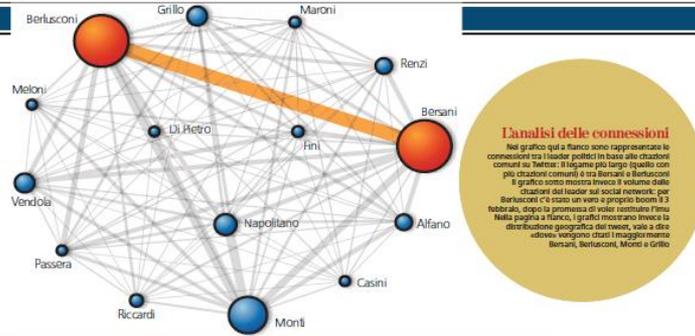
Berlusconi infiamma Twitter dopo la proposta sull'Imu

GRANDE REPORT
SIGILLI DELLA PRIMA PAGINA

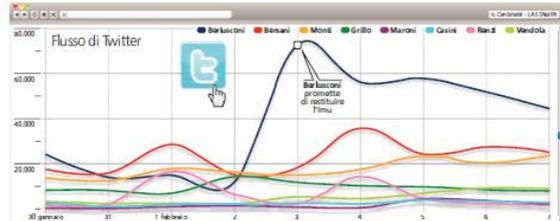
La stampa avrà da oggi i propri dati di riferimento e di analisi. Il progetto Tycho-La Stampa, che prevede la creazione di un database di dati di Big Data, è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia. Il progetto Tycho-La Stampa è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia.

Il progetto Tycho-La Stampa è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia. Il progetto Tycho-La Stampa è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia.

Il progetto Tycho-La Stampa è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia. Il progetto Tycho-La Stampa è un progetto che ha come obiettivo quello di analizzare i dati di Big Data di questa prima pagina pubblica, la quale prima parlava di politica e di economia, e ora di politica e di economia.



L'analisi delle connessioni
Nel grafico qui a fianco sono rappresentate le connessioni tra i leader politici in base alle citazioni contenute nei tweet. Il grafico più grande (quello con più citazioni comuni) è tra Berlusconi e Berlusconi. Le citazioni dei leader sul social network per Berlusconi ci sono un voto a proprio favore (1,3), ma, dopo la promessa di voler restituire l'Imu nella pagina a fianco, i grafi mostrano invece la distribuzione geografica dei tweet, vale a dire dove vengono citati maggiormente i nomi Berlusconi, Berlusconi, Monti e Grillo.



grillo e Stella. Molto chiaro in realtà, così, è il grafico che mostra la distribuzione geografica dei tweet, vale a dire dove vengono citati maggiormente i nomi Berlusconi, Berlusconi, Monti e Grillo.

messaggio elettorale
comunicato Giovanni Maria Flick

**VOGLIO UN'ITALIA
CHE ROTTAMI I FURBI
E SCELGA I BRAVI.**

Giovanni Maria Flick
l'uomo giusto per regole giuste.

Capolista al Senato, in Piemonte e Lazio

**CENTRO
DEMOCRATICO**

CON BRUNO TABACCHI

**Bersani
Presidente**

Intervista

Marco Bardazzi

Non sono gli stessi sondaggi politici. Non sono neppure accettati e accreditati dall'opinione pubblica. In questi giorni si ripresentano ogni giorno su Internet i grafici di questo genere e quasi ininterrottamente si ripresentano le previsioni quattro settimane (prima e dopo il voto) sull'attuale possibilità di vittoria di una lista di dati di voto e di voto. In questi giorni si ripresentano ogni giorno su Internet i grafici di questo genere e quasi ininterrottamente si ripresentano le previsioni quattro settimane (prima e dopo il voto) sull'attuale possibilità di vittoria di una lista di dati di voto e di voto.

**"Incrociando tutti quei dati
potremmo scoprire
anche chi ama la cioccolata"**

Il fisico Chessa: fondamentale analizzare le relazioni

Alessandro Chessa

Il fisico Chessa: fondamentale analizzare le relazioni. Il fisico Chessa: fondamentale analizzare le relazioni. Il fisico Chessa: fondamentale analizzare le relazioni.

Tycho La Stampa 18-02-2013:

Da oggi con La Stampa il primo dei due volumi *

LA STAMPA

QUOTIDIANO FONDATA NEL 1847

LUNEDÌ 18 FEBBRAIO 2013 - ANNO LVII N. 48 - 1,20 € IN TUTTI I PUNTI VENDITORI SU TUTTO IL TERRITORIO NAZIONALE - D.L. 33/03/01 ART. 1 COMMA 1, LETT. A) - WWW.LASTAMPA.IT

Roma si prepara all'accoglienza dei pellegrini

Il Papa che se ne va commuove i fedeli "Pregate per me"

Il messaggero non usare Dio per il potere



Folla festosa in piazza S. Pietro per l'Angelus di Benedetto XVI. Galassi, Martini e Tosielli DA FERRARIS/A3/AGF/13

TREDECI MINUTI E UN SOLO UFFICIO "RIMANI, RIMANI"

Incasso alla velocità della luce

Quando, a mezzogiorno, un pasticcino bianco appare sulla piazza San Pietro, un grido si leva più forte degli altri: «Rimani, rimani». Sono quelli di Milla Chiesa. CONTINUA A PAGINA 117

PARLANO I CARDINALI Vingt-Trois (Parigi) "Ora dobbiamo aprire il mondo"

Alberto Martini A3/AGF/13

Dolan (New York) "Favorito? Chi lo dice fuma marijuana"

WITTENBERG/AGF/13

Berlusconi a Torino: il comico, un pericolo per la democrazia. Mi ubriaco se il premier e Casini non saranno eletti

Elezioni, la guerra della tv

Grillo disdice all'ultimo un'intervista, Monti chiede il confronto a tre Milano: folla al comizio Pd, a sorpresa appare Prodi: resteremo uniti

IL MONOPOLISTA DELLE PIAZZE HA GIÀ VINTO

ELISABETTA GIULIANO

REPORTAGE

Amici e vicini "Così Pistorius è impazzito"



Pistorius e la fidanzata ucraina CONTINUA A PAGINA 116

Grillo ha deciso di non apparire in tv perché non ne ha bisogno. Poi, perentoriamente, si è recato alle battaglie del piccolo schermo e rimanere fedele alla strategia delle piazze, di cui è stato un frequentatore quasi monopolista, perché ha già vinto.

CONTINUA A PAGINA 114

IL PRIMO PASSO DI ROMANO VERSO IL COLLE

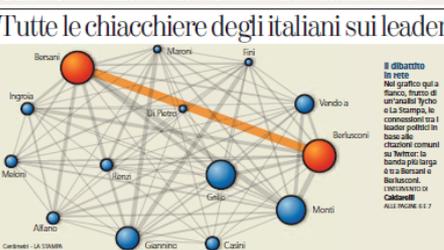
FABIO MARTINI

Con un numero a sorpresa, si è aperta la piazza Duomo a Milano lunedì notte con il debutto nella storia della Repubblica. Si è aperta nel mondo più tranquillo, con i carabinieri.

CONTINUA A PAGINA 115

Tutte le chiacchiere degli italiani sui leader

ESCLUSIVO: LA MAPPA DELLE CONVERSAZIONI POLITICHE CITTÀ PER CITTÀ SUL FLUSSO DI TWITTER



Il dibattito in rete

Nel grafico qui a fianco, frutto di un'analisi Tycho e La Stampa, le connessioni tra i leader politici in base alle chiacchiere comuni su Twitter. La banda più larga è tra Berlusconi e Monti.

CONTINUA A PAGINA 117

IL BIPOLARISMO È DIVENTATO UN POKER

GIANNI RIOTTA

Il Partito democratico guidato da Pier Luigi Bersani resta, se non gli uomini e i giuristi del web in Italia, favorito per la vittoria elettorale, domenica ventura.

CONTINUA A PAGINA 117

LA STAMPA

99,9% FARMACUTI 01 LA SALUTE DEGLI ITALIANI DAL 1925 99,9% ARMACUTI 02

NUOVA TENDENZA IN AMERICA: BASTA MEETING, PIÙ TEMPO PER RAGGIUNGERE GLI OBIETTIVI

Troppe riunioni rovinano il lavoro

FABIO MARTINI/AGF/13

chi critica la sua proposta per risolvere i problemi dell'immigrazione, John Maynard Keynes rispondendo che così lungo perché ancora tutti morti. Forse gli editori del New York Times avevano in mente questa battuta, quando hanno deciso di lanciare la loro rivista nei confronti di troppi meeting, le riunioni, gli incontri che affollano le nostre giornate di lavoro, togliendoci la sp...

Il rubino di ogni settimana

Elavast, come apprezzare il modo del lavoro italiano

Mauro Tomareo A3/AGF/13

CONTINUA A PAGINA 112

CLINIAN'S

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

ABBONAMENTI | ARCHIVIO | PIÙ VISTI | SOCIAL | METEO | TUTTOAFFARI | LAVORO | LEGALI | NECROLOGIE | SERVIZI | 

LA STAMPA.it

Cerca... 

TORINO - CUNEO - AOSTA - ASTI - NOVARA - VCO - VERCELLI - BIELLA - ALESSANDRIA - SAVONA - IMPERIA e SANREMO

VOCI DI MILANO

ATTUALITÀ | OPINIONI | ECONOMIA | SPORT | TORINO | CULTURA | SPETTACOLI | MOTORI | DONNA | CUCINA | SALUTE | VIAGGI | EXTR@ | FOTO | VIDEO

HOME | POLITICA | ESTERI | CRONACHE | COSTUME | TECNOLOGIA | SCIENZA | AMBIENTE | LAZAMPA | I TUOI DIRITTI | MARE | MONTAGNA | SPECIALI

Voto, le parole che decideranno la volata Big data - Il bipolarismo diventa un poker

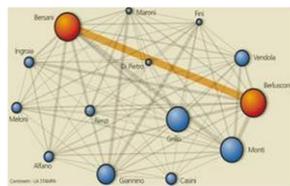
GIANNI RIOTTA

Di chi si parla su twitter? Dove? E di cosa? In esclusiva la mappa delle conversazioni politiche città per città. Sei regioni in bilico

Ecco tutte le chiacchiere degli italiani sui leader nell'analisi Tycho per La Stampa. Così Berlusconi resta sempre al centro del dibattito anche in rete.

LO SPECIALE ELEZIONI / LE VOCI DELLA POLITICA SU INTERNET

APP LASTAMPA.IT Trova il candidato / + Segnala i tuoi video

«Mps a fine 2011 fuori controllo»

Bankitalia accusa i vecchi vertici

Galileo Pizzoli A3/AGF/13

Yoani: il viaggio della libertà

La blogger cubana in volo per il Brasile

Yoani Sánchez A3/AGF/13

CLINIAN'S

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

Da paracaduto in America si è presentata a miglior vita. Palla sta che il giornale della città più fredda al mondo ha lanciato proprio questa campagna, consigliando a tutti di eliminare o accorciare i meeting, per concentrarsi invece sulle attività che producono davvero qualcosa di utile.

Tycho La Stampa 18-02-2013:

6 | Primo Piano | L'ESPRESSO

-6 giorni al voto

Sei regioni in bilico Berlusconi è sempre al centro del dibattito

Analisi di Tychobigdata-Imt dei tweet sui leader Guida il Cavaliere, ma i commenti sono spesso negativi

GIORGIO CALZAVALLI

I dati di Tychobigdata-Imt presentati oggi a volume di twitter di alcuni dei principali politici in corsa per le prossime elezioni. Individuando che un tweet ritrasmesso a un messaggio di 140 caratteri in cui si possono indicare degli argomenti con un termine conosciuto da un hashtag o un nostro caso abbiamo considerato tutti i messaggi inviati con hashtag dei candidati principali nelle province studiate, nonché altre personalità che generano un volume consistente di dati. Una domanda più sottile è di un nuovo modo presentato per Silvio Berlusconi l'Espresso, data dall'istituto delle prossime elezioni di Papa Benedetto XVI. Da notare che in quel giorno molti politici sono ancora al hashtag #Papa o simili. Per Silvio



Docente di fisica teorica

Guido Calzavalli è docente di Fisica Teorica, docente presso l'Istituto di Matematica e Fisica di Lucca, e ricercatore al Mathematical Sciences di Londra.

Berlusconi di sono circa 8000 tweet di quasi 12000, quasi il triplo (dopo il 2008) del tweet in cui sono presentati Mario Monti e il Papa.
Il problema che pone, se si vuole andare oltre l'analisi numerica di questo fenomeno, è di comprendere quanto questi volumi di messaggi siano correlati ai candidati, ovvero se critica o anche semplicemente neutra. Un primo modo di analizzare le informazioni è quello di analizzare le occorrenze dei vari hashtag nel tweet. Questa è la scelta che presentiamo in questa analisi. Basandoci a ruota dei politici da analizzare, che a questo proposito in Italia dei principali argomenti cui sono associati. A breve presenteremo un'analisi di questo contenuto di hashtag. Finito con l'analisi di questo fenomeno di dati, in un prossimo numero di Tychobigdata-Imt.



Il progetto Tycho-La Stampa

Da oggi all'indietro voto2013.itastampapresentaanalisiemappediinteressi del progetto Tycho-La Stampa. L'obiettivo è di fornire informazioni sul voto del nostro giornale, attraverso dati provenienti da twitter e commentari risultati del progetto. La analisi Tycho-La Stampa di questo progetto sono state inviate ai ricercatori dell'Istituto di Lucca, analisti, fisici, statistici, informatici, economisti, e @tychobigdata.

Le parole e le priorità che decideranno la volata

Di chi si parla su twitter? Dove? E di cosa? Ecco le risposte

MIXA-CONTO.

LA FLESSIBILITÀ DI UN CONTO CORRENTE E IL RENDIMENTO SICURO DI UN DEPOSITO AL 4%.

INGREDIENTI

- Conto Corrente a zero costi di apertura con Interest
- Conto Corrente a zero costi di apertura con Interest
- Parzialmente di deposito a zero costi di apertura con Interest
- o a 12 mesi
- Parzialmente di deposito a zero costi di apertura con Interest
- o a 12 mesi
- Parzialmente di deposito a zero costi di apertura con Interest
- o a 12 mesi

PREPARAZIONE

Aperto in un solo click, mantenete il vostro conto corrente a zero costi di apertura con Interest. Il conto corrente a zero costi di apertura con Interest è aperto in un solo click, mantenete il vostro conto corrente a zero costi di apertura con Interest. Il conto corrente a zero costi di apertura con Interest è aperto in un solo click, mantenete il vostro conto corrente a zero costi di apertura con Interest.

Maggiori informazioni sulla pagina IMI, chiamando il numero verde 190 100 100 o visitando www.bpm.it

BPM BANCA POPOLARE DI MIANO
Il futuro è di chi fa.

Giorgio Ricca

La rete non accennava sulla sua completa fine, ma con gli aiuti di Monti, gli ex presidenti della Camera Carlo e Prodi, la rete si è riattivata. Monti non contraria, ma dice di esserne felice. Monti non contraria, ma dice di esserne felice. Monti non contraria, ma dice di esserne felice.



Giustizia la corsa nel giorno storico della discesa annunciata di papa il nostro XVI. Berlusconi è stato il punto di partenza per il nostro XVI. Berlusconi è stato il punto di partenza per il nostro XVI. Berlusconi è stato il punto di partenza per il nostro XVI.

preparati. Nel 2013 non abbiamo visto una svolta italiana medievale. Altre correnti di pensiero si sono presentate al leader della Lega Nord. Berlusconi è stato il punto di partenza per il nostro XVI. Berlusconi è stato il punto di partenza per il nostro XVI.

Data flow



Stream demon
Search cron

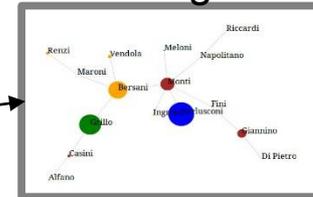


**CN Analytics
Geocator**

WebServices



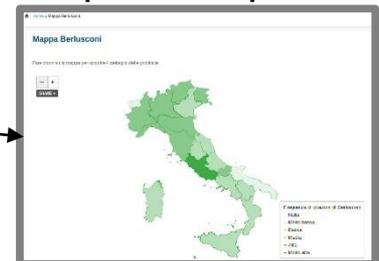
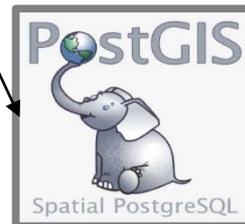
Web widget



Tycho Drupal



Mapbox maps



Progetto INSIDER:

Le tecniche della **teoria delle reti** permettono di elaborare e sintetizzare grandi quantità di dati al fine di far risaltare quelle **relazioni deboli** che non sono immediatamente visibili ma importanti ai fini della legalità.

Persone e società
formano delle reti di
relazioni che possono
essere analizzate
mediante la **teoria delle
reti**

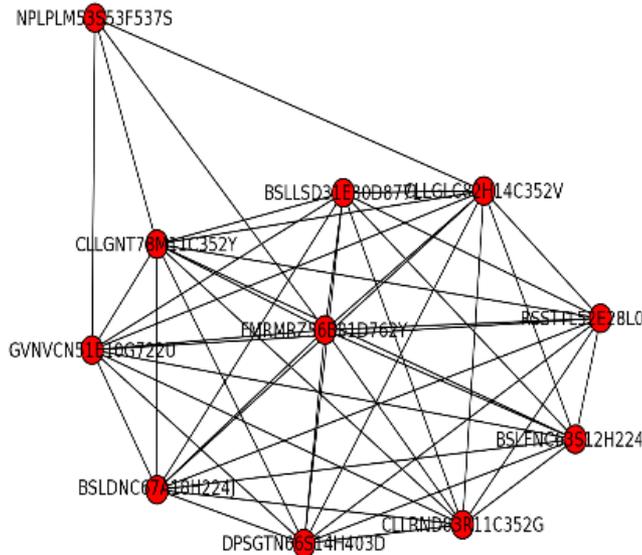


1. Analisi delle visure camerali delle province di Vibo Valenza e Crotona
2. Identificazione dei legami reali e potenziali tra i vari soggetti economici (persone ed aziende)
3. Costruzione delle reti e visualizzazione interattiva
4. Calcolo di indicatori di **centralità** (importanza strategica del soggetto economico all'interno della rete di relazioni) e di **aggregazione/comunità** (identificazione di gruppi di soggetti economici legati da legami di diversa natura)
5. Sviluppo di widget che aiutano la visualizzazione e l'identificazione dei soggetti più da attenzionare.

Analisi Preliminari:

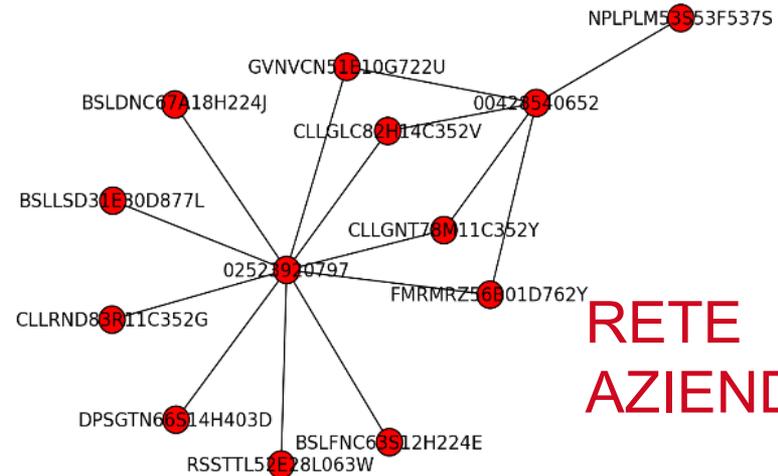
1. Province di Vibo Valenza e Crotona
2. Settori merceologici ATECO:
 41. Costruzione di edifici
 42. Ingegneria civile
 43. Lavori di costruzione specializzati
1. Identificazione di due network principali:
 - Aziende-soci
 - Aziende-amministratori
1. Analisi della rete bipartita e proiezione su aziende, soci e amministratori
2. Analisi della rete delle aziende, dei soci e delle persone

Rete Aziende Soci



RETE SOCI

Quali sono i gruppi di persone che detengono quote, direttamente o indirettamente, nelle stesse società? E chi fa parte di ciascun gruppo?



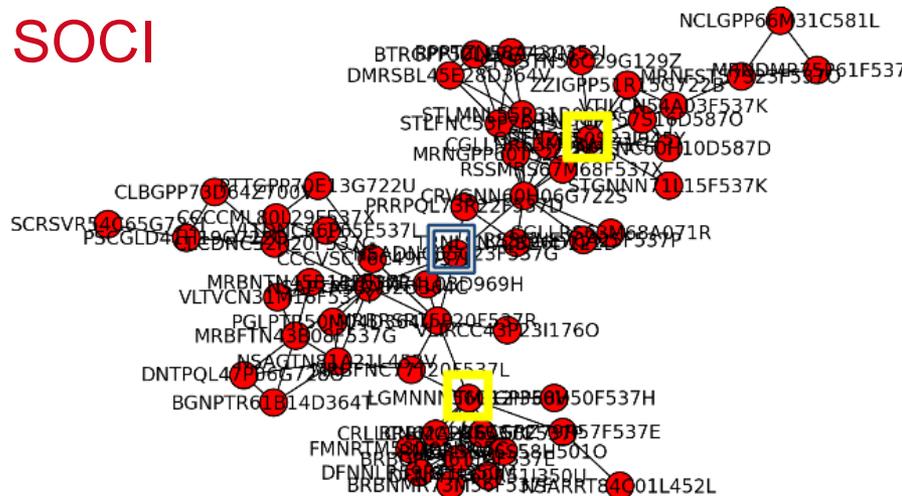
RETE
AZIENDE

Quali sono le aziende che sono connesse tramite azionisti comuni? Quali sono le caratteristiche di queste aziende?

Rete Soci

- Alcune reti posso essere di **grandi dimensioni**, difficili da interpretare.
- Necessità di introdurre **indicatori** che evidenzino i nodi (società o persone) che ricoprono posizioni particolari all'interno del network.
- Possibilità di effettuare visualizzazioni dinamiche che sintetizzano le principali caratteristiche (**widgets**)

RETE SOCIE



Il più numeroso
gruppo di aziende
legate tra di loro da
azionisti in comune è
formato da 56 nodi

Indicatori di centralità

Gli indicatori di **centralità** servono ad identificare quei nodi che hanno un ruolo di particolare importanza all'interno della rete sociale e commerciale, poiché sono titolari di numerose connessioni, che permettono di:

- ***Ottenere informazioni e risorse da più fonti contemporaneamente sfruttando la rete e quindi anche più velocemente rispetto ad altri metodi***
- ***Avere il ruolo di broker di informazioni e risorse***

L'importanza di questi metodi è legata alla totale "imparzialità" dell'analisi che non fa uso di informazione pregressa ma la deriva direttamente dai dati delle camere di commercio.

Rete delle imprese di costruzioni

Un esempio legato alla **rete delle imprese di costruzioni** della provincia di Vibo.

La rete dei soci delle varie società (che possono avere quote in altre società) viene analizzata per formare la rete bipartita (relazioni persone-aziende) e quindi *proiettata* per le persone o per le aziende.

Studiamo le misure di centralità senza inferire ulteriori informazioni sulle persone-aziende coinvolte.

Concentrandoci sulle persone che sono più centrali (hanno una *betweenness* più elevata, cioè una maggiore capacità di stabilire relazioni economiche tra società differenti), da una breve verifica sulla stampa si scopre che...

Crime detection

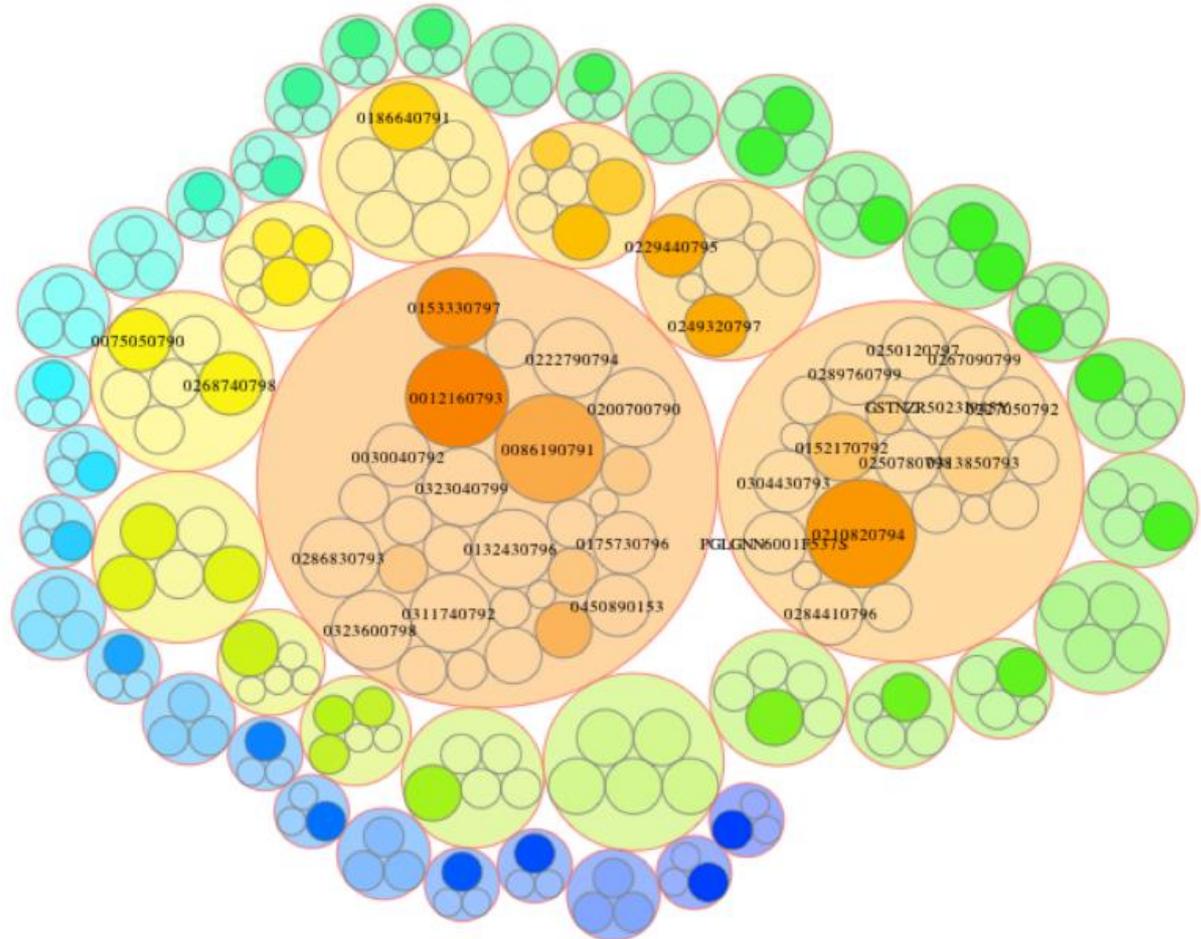
Potenzialità dell'analisi di rete come indicatori di criminalità

- legami criminali risultano spesso in legami di affari
- le partecipazioni tra aziende e i soci in comune sono portatori di pratiche criminali (e non) che si propagano alle imprese associate
- l'analisi di rete cattura i legami prevalenti tra gli imprenditori e tra i soci e li fa risaltare in maniera "automatica" attraverso semplici misure di centralità
- in maniera analitica è poi possibile ricostruire con questi metodi le società collegate anche da legami deboli quali l'appartenenza ad una stessa comunità (per esempio un piccolo paese).

Visualizzazione

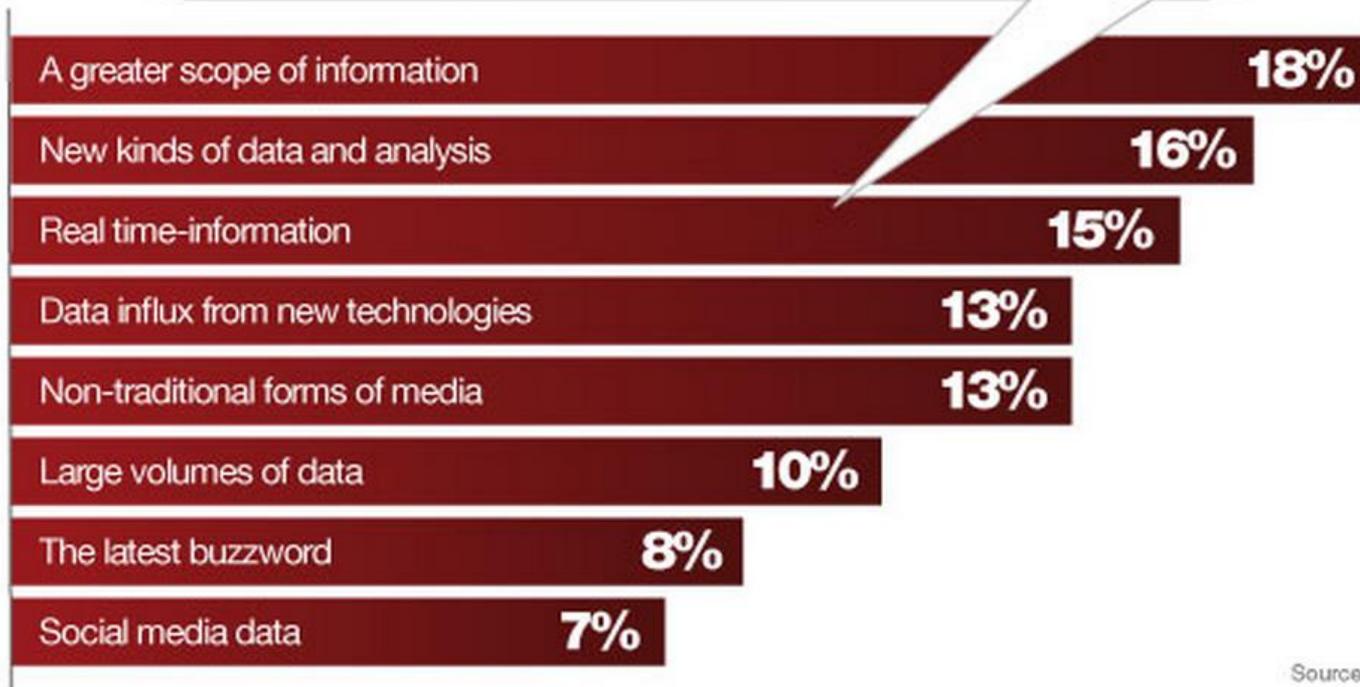
Clustering e Betweenness

- Rete delle imprese
- Le aziende sono divise in gruppi o comunità di connessioni
- La grandezza del nodo corrisponde al grado (n. di connessioni)
- Il colore del nodo corrisponde alla betweenness (centralità)

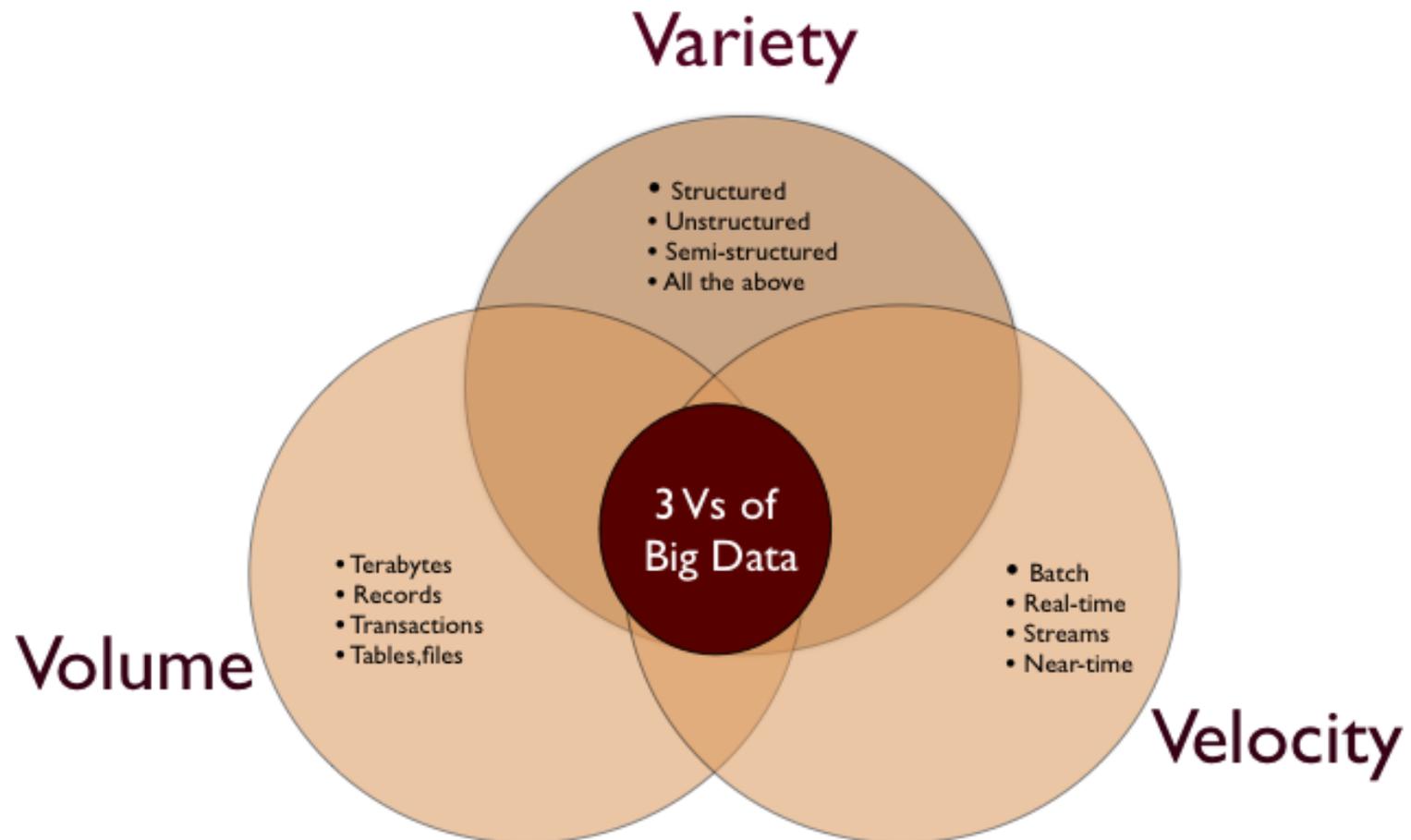


Cosa sono i Big Data?:

What do business executives think “big data” is?



Le 3 V dei Big Data:



Fatti importanti sui Big Data :

- 75% of companies say they will increase investments in Big Data within the next year (source: Avanade)
- 70% of data is created by individuals – but enterprises are responsible for storing and managing 80% of it (source: CSC)
- 59% of organizations lack the tools required to manage data from their IT systems (source: Saffron Technologies)
- 90% of the world's total data has been created just within the past two years (source: IBM)
- According to new research by Capco, 62% of banks believe that managing and analyzing big data is important to their success. However, only 29% say they are currently extracting enough commercial value from data.
-

Big Data nei servizi finanziari :

-“Analytics, customer-centricity, and multichannel technologies are major trends across the market,”

-“Banks increasingly want a core system that utilizes data analytics in order to provide a more complete view of the customer, which then allows for better customer-bank communication,” Narter writes.

Bart Narter, SVP di Celent, e co-autore del report, Core Banking Solutions for Midsize Banks: A Global Perspective.

Principali Azioni nel settore Bancario:

>Consolidamento dei **dati interni**, come ad esempio le operazioni di addebito e di credito, storie di acquisto, l'utilizzo dei canali di comunicazione, preferenze, fidelizzazione, ecc

>Espansione dei dati interni strutturali per includere la raccolta e l'analisi di dati provenienti da **fonti esterne**, tra cui click stream web, interazioni sociali (Facebook e Twitter), informazioni di geo-localizzazione.

Potenziali benefici dei Big Data:

Risk Management/Fraud Detection

Con l'aggiunta **dispositivi alternativi** per le transazioni c'è la possibilità di monitorare i cambiamenti nel comportamento al di là di ciò che sta accadendo sul conto corrente di un cliente: le banche possono isolare nuove frodi o trigger di rischio e intervenire tempestivamente se necessario

Targeting

Comprendere meglio i consumatori, e quindi poter offrire a un cliente o prospect al momento giusto, permette un istituto finanziario di ottimizzare la gestione delle relazioni con i clienti redditizi, a lungo termine. L'aggiunta di una grande quantità di dati online esterni non strutturati amplia di molto la conoscenza del cliente, potenzialmente migliorando sia l'efficacia e l'efficienza degli sforzi di marketing..

Un approccio graduale ai Big Data nelle banche:

La **sfida** per le banche è quella di andare oltre le tecnologie *legacy*, che prevedono 'silos' di dati aziendali e scarsa dimestichezza con gli *analytics* necessari per maneggiare grandi quantità di dati non strutturati.

Un possibile approccio 'step by step':



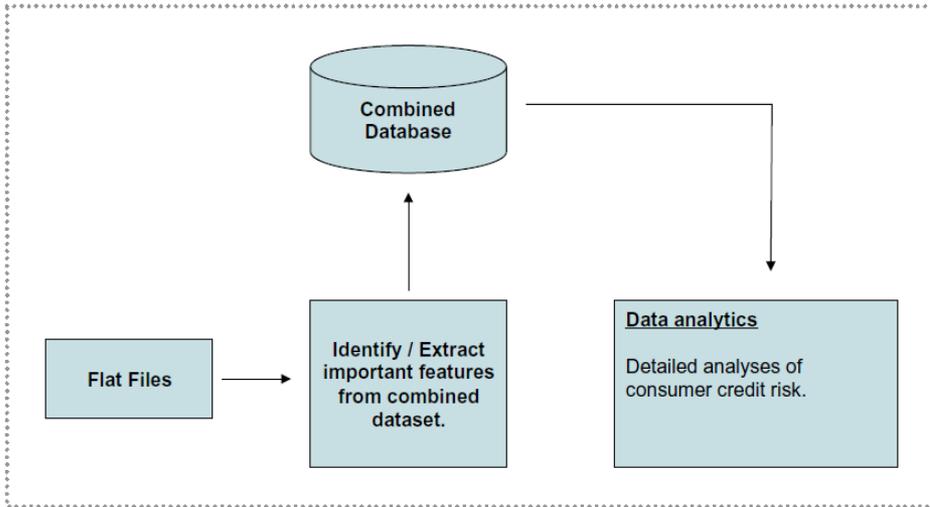
Source: Novantas, LLC

More data doesn't fix bad analytics!

Caso studio: Consumer credit-risk models via machine-learning algorithms (Khandani et al., JBF)

- Applicazione di tecniche di machine-learning per costruire modelli predittivi per il rischio di credito commerciale.
- Combina dati su transazioni e dei credit bureau e migliora capacità produttiva su delinquencies e defaults su carta di credito con un risparmio tra il 6 e il 25% delle perdite attese
- Implicazioni per l'analisi predittiva microfondata del rischio sistemico

Caso studio: Integrazione dati eterogenei



Individual level	Account level
Credit score	Age of Account
File age	Open/closed flag and date of closure
Bankruptcy (date and code)	Type (CC, MTG, AUT, etc.)
MSA and zip code	Balance
	Limit if applicable
	Payment status
	48-Month payment status history

Deposit data	
Checking account balance	CD account balance
Brokerage account balance	IRA account balance
Savings account balance	

Transaction data

Transaction count
Total inflow
Total outflow

By Channel:

ACH (count, inflow and outflow)
ATM (count, inflow and outflow)
BPY (count, inflow and outflow)
CC (count, inflow and outflow)
DC (count, inflow and outflow)
INT (count, inflow and outflow)
WIR (count, inflow and outflow)

By Category

Employment inflow
Mortgage payment
Credit-card payment
Auto loan payment
Student loan payment
All other types of loan payment
Other line of credit payment
Brokerage net flow
Dividends net flow
Utilities payment
TV
Phone
Internet

By category (cont.)

Hotel expenses
Travel expenses
Recreation
Department store expenses
Retail store expenses
Clothing expenses
Discount store expenses
Big box store expenses
Education expenses
Total food expenses
Grocery expenses
Restaurant expenses

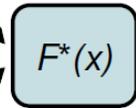
Bar expenses
Fast food expenses
Total rest/bars/fast-food
Healthcare related expenses
Fitness expenses
Health insurance
Gas stations expenses
Vehicle expenses
Car and other insurance
Drug stores expenses
Government
Treasury
Pension inflow
Collection agencies
Unemployment inflow

Machine learning: il modello CART

Inputs describing consumer j

consumer level categorical
expenditures

Forecast Model



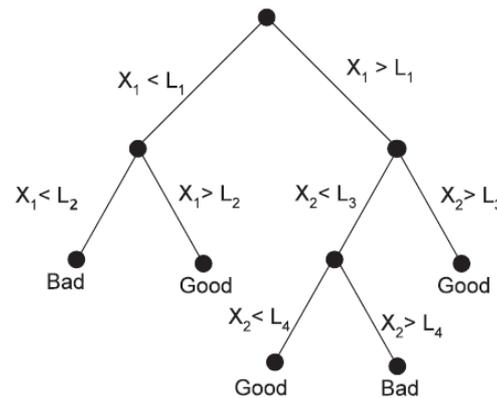
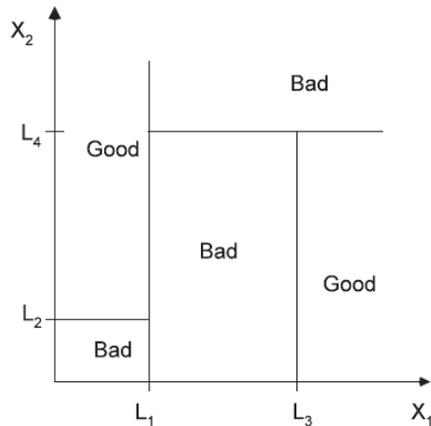
consumer credit history &
financial behaviors

Credit Risk Forecast of consumer j

$P(X_j)$: probability of consumer j
becoming 90+ days delinquent
within next 3 months

Generalized classification and regression trees

Pruning and boosting



Risultati: importanti implicazioni per credit line e macroprudential risk management

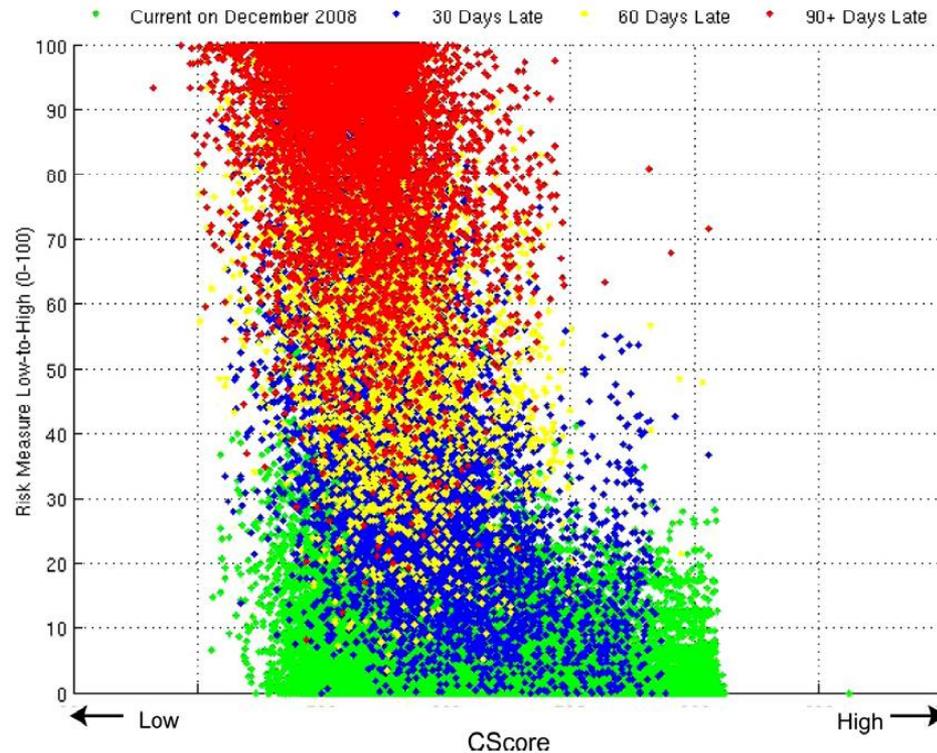


Fig. 13. Color-coded comparison of December 2008 CScore (x -axis, and where higher values are associated with lower credit risk) and machine-learning forecasts of 90-days-or-more delinquency rates over subsequent 3-month windows using realized delinquency events from October to December 2008 and September 2008 feature vectors. Once calibrated, the machine-learning model is applied to the December 2008 feature vector, and the resulting "fitted values" are plotted (y -axis) against the December 2008 CScores. The color coding indicates whether an account is current (green), 30-days delinquent (blue), 60-days delinquent (yellow), or 90-days-or-more delinquent as of December 2008.