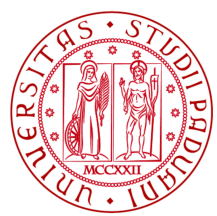
La statistica per il web

Dalle ricerche su google agli algoritmi di traduzione, dall'analisi delle reti di contatti alla sentiment analysis sui social media

Finos Livio

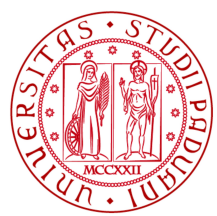
Dipartimento di Scienze Statistiche
Università degli Studi di Padova

Summer School
La matematica incontra le altre Scienze
San Pellegrino Terme, 8-9-10 Settembre 2014



Come pollicino nel bosco...

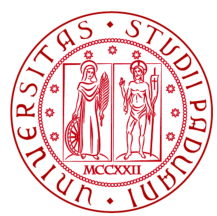
Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio



Come pollicino nel bosco...

Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio

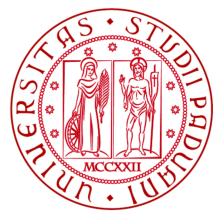
- quando 'postiamo' su FaceBook o su Twitter



Come pollicino nel bosco...

Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio

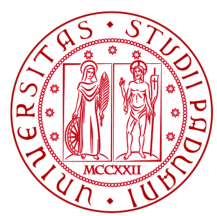
- quando 'postiamo' su FaceBook o su Twitter
- facciamo una ricerca su un motore di ricerca



Come pollicino nel bosco...

Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio

- quando 'postiamo' su FaceBook o su Twitter
- facciamo una ricerca su un motore di ricerca
- o semplicemente scriviamo una e-mail

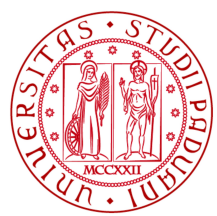


Come pollicino nel bosco...

Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio

- quando 'postiamo' su FaceBook o su Twitter
- facciamo una ricerca su un motore di ricerca
- o semplicemente scriviamo una e-mail

Internet produce ogni giorno migliaia di terabyte (1000 gigabyte) di dati.

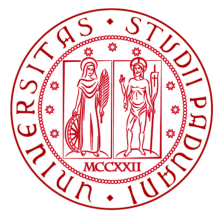


Come pollicino nel bosco...

Quando navighiamo su internet lasciamo molte informazioni.
Ad esempio

- quando 'postiamo' su FaceBook o su Twitter
- facciamo una ricerca su un motore di ricerca
- o semplicemente scriviamo una e-mail

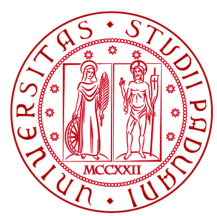
Internet produce ogni giorno migliaia di terabyte (1000 gigabyte) di dati.
Qualcuno li usa? o vengono buttati via? **servono**?
Ci possono aiutare a **ritrovare la via**?



lo statistico: il lavoro più sexy del decennio :)

The sexiest job in the next 10 years will be statisticians.
(H. Varian, Chief Economist, Google, 2010)

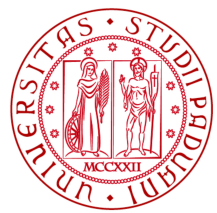
perché dice questo?



La statistica serve a quasi tutte le discipline scientifiche...

come ad esempio:

- fisica (anche nella scoperta del bosone di Higgs!)
- sociologia e fenomeni sociali e marketing
- produzione industriale
- previsioni politiche (sondaggi elettorali, exit-poll)
- analisi economiche (previsione degli effetti della crisi)
- sport (nel basket già molto diffuso, scommesse)
- psicologia e neuroscienza
- biologia e medicina



La statistica serve a quasi tutte le discipline scientifiche...

come ad esempio:

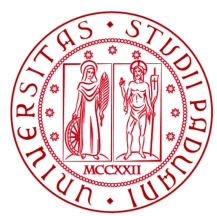
- fisica (anche nella scoperta del bosone di Higgs!)
- sociologia e fenomeni sociali e marketing
- produzione industriale
- previsioni politiche (sondaggi elettorali, exit-poll)
- analisi economiche (previsione degli effetti della crisi)
- sport (nel basket già molto diffuso, scommesse)
- psicologia e neuroscienza
- biologia e medicina

da quando Google (e gli altri) ha risolto il problema di raccogliere e rendere accessibile i dati prodotti, si è accorta di aver bisogno di qualcuno che li 'leggesse'!

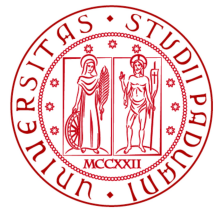
La statistica ora è affamata di dati della rete

Lo statistico raccoglie e rielabora i dati, li ascolta, li analizza con gli **strumenti matematici** e racconta agli altri cosa hanno da dirci...
In questo talk vedremo alcuni esempi (nessuna formula, solo idee)

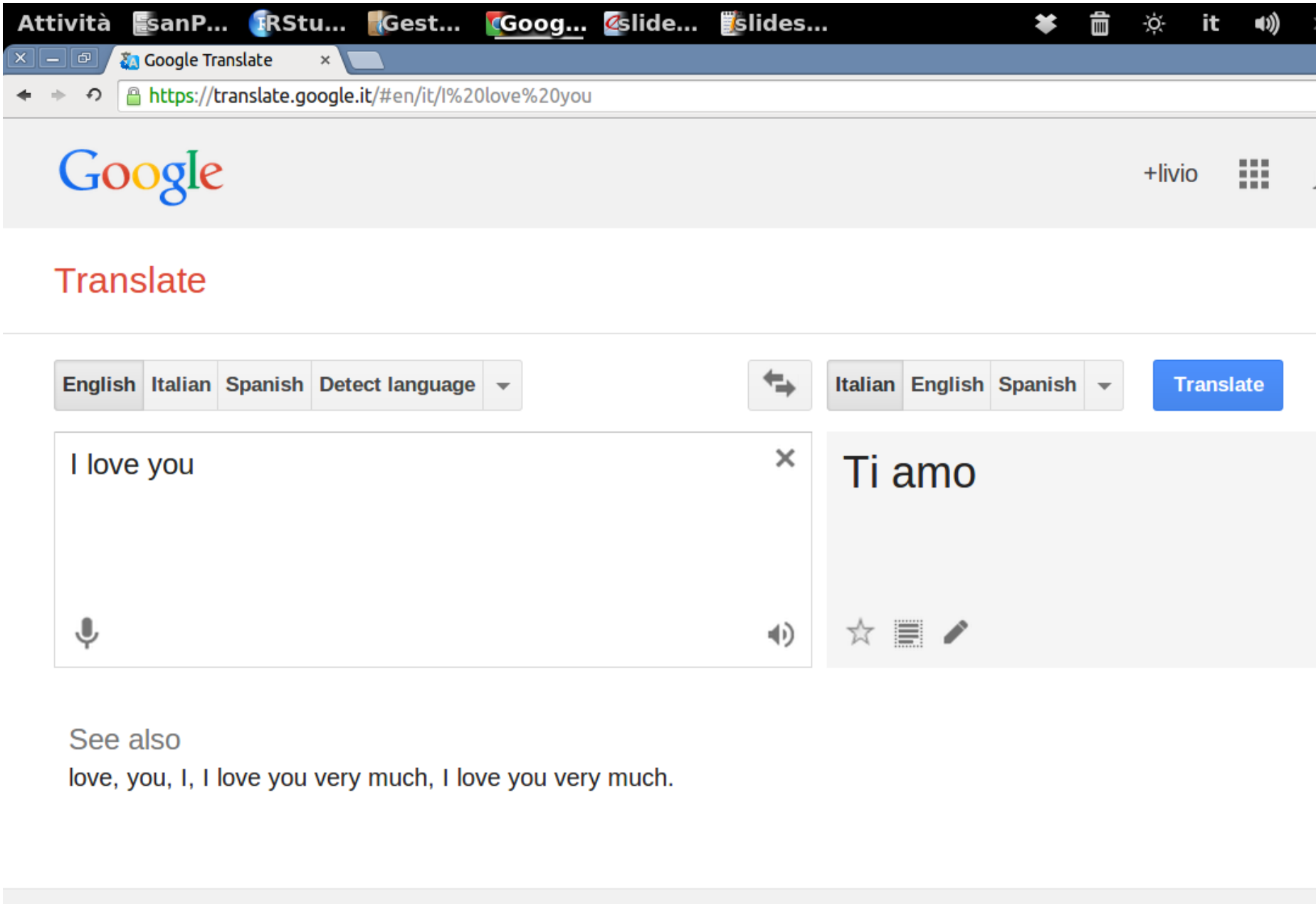
- Google Translate (sistema di traduzione automatica)
- Google Trend (ricerche su motore di ricerca)
- Sentiment/Content analysis
- Social network analysis (su FaceBook)



Google Translate



Google Translate



Come funziona

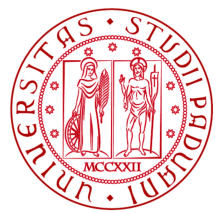
Traduzione parola per parola + eccezioni e forme particolari?
NO: chi ci ha provato ha fallito
nel nostro esempio:
I love you: [io] [amore/amo/ami/amiamo/..?] [tu/ti/te/voi/vi..?]

perché: molte, moltissime le eccezioni
difficoltà di gestire le nuove parole, forme verbali e slang
lavoro immane, soprattutto considerando le 80 lingue di Translate.

Come funziona

Statistical Machine Translation!

- prende le migliaia di siti in doppia lingua, i libri tradotti etc. (es. IT-EN)
- - per ogni parola conta quante volte la parola inglese è stata tradotta in italiano. es:
    I →[99% io]
    love → [60% amore/ 20% volere bene/ 20% provare piacere]
    you → [20% tu/ 20%ti/ 20% voi/ 20% vi]
  - lo fa anche per pezzi di frase più lunghi. es:
    'I love you' → ['ti amo' 50%/ 'ti voglio bene' 20%/ 'vi amo' 10%/ 'vi voglio bene' 10%/ etc]
- propone la traduzione più frequente, che più probabilmente è quella che cercate voi!



Google Translate: parola you

Definitions of you

pronoun

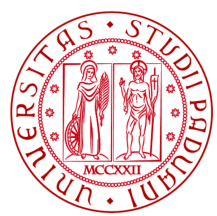
used to refer to the person or people that the speaker is addressing.
"are you listening?"

used to refer to any person in general.
"after a while, you get used to it"

Translations of you

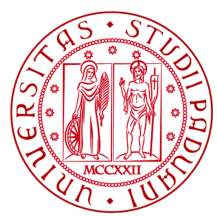
pronoun

voi	you, ye
vi	you, yourself, ye
ti	you, thee, yourself, thyself, ye
te	you, yourself, thee, ye
tu	you, thou, ye
lei	she, you, her, herself
la	it, her, you
le	them, you, it, ye
ve	you, ye, yourself
loro	their, them, they, theirs, ye, you

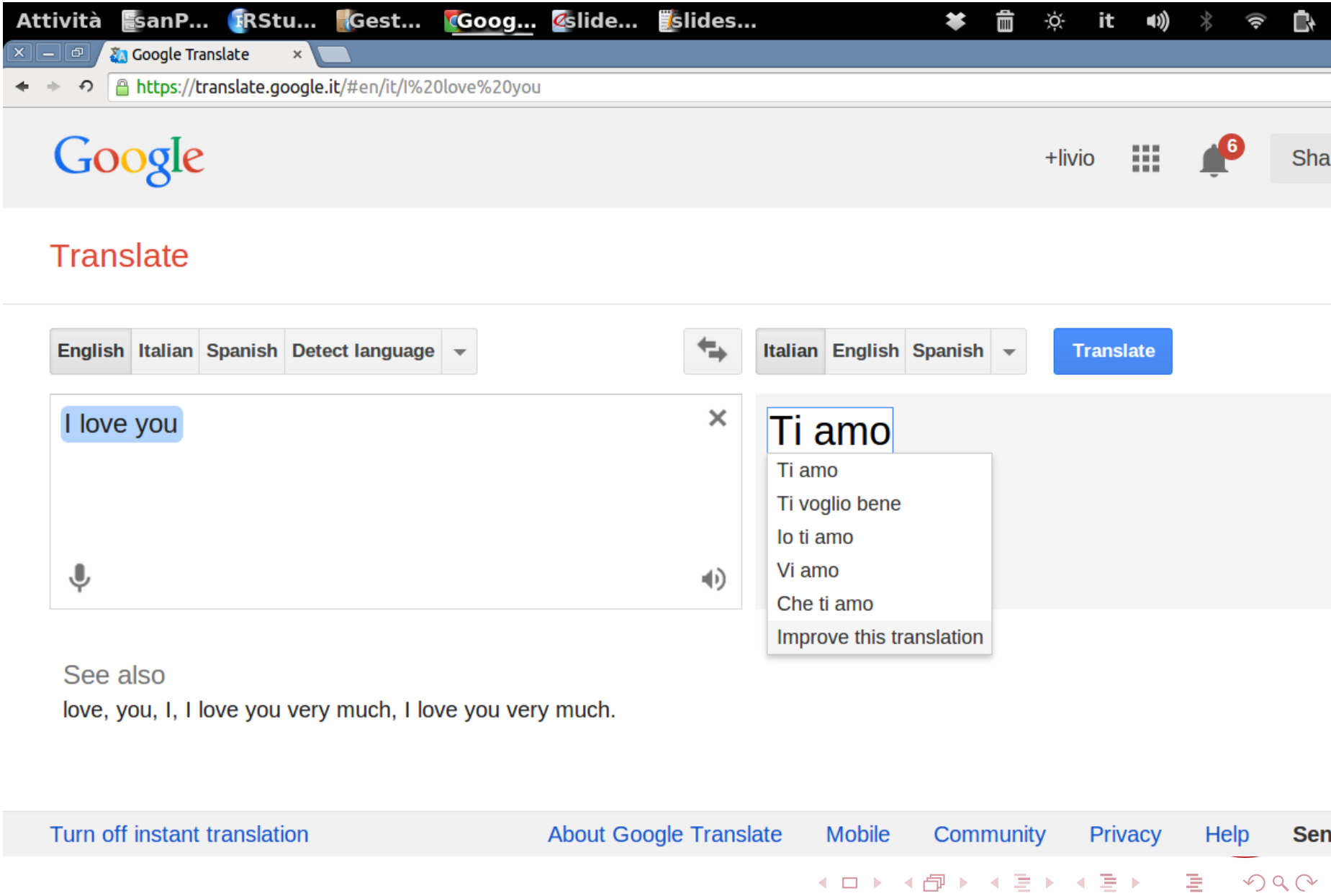


Google Translate: probabilità e non certezze

Studiare le frequenze delle traduzioni della parola **You** equivale a stimare la **probabilità** che la parola venga tradotta in un certo modo.
Traduciamo la parola o la frase nel modo che stimiamo essere più verosimile (principio di **Massima Verosimiglianza**)

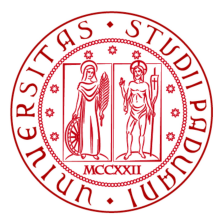


Google Translate: intera frase I love you



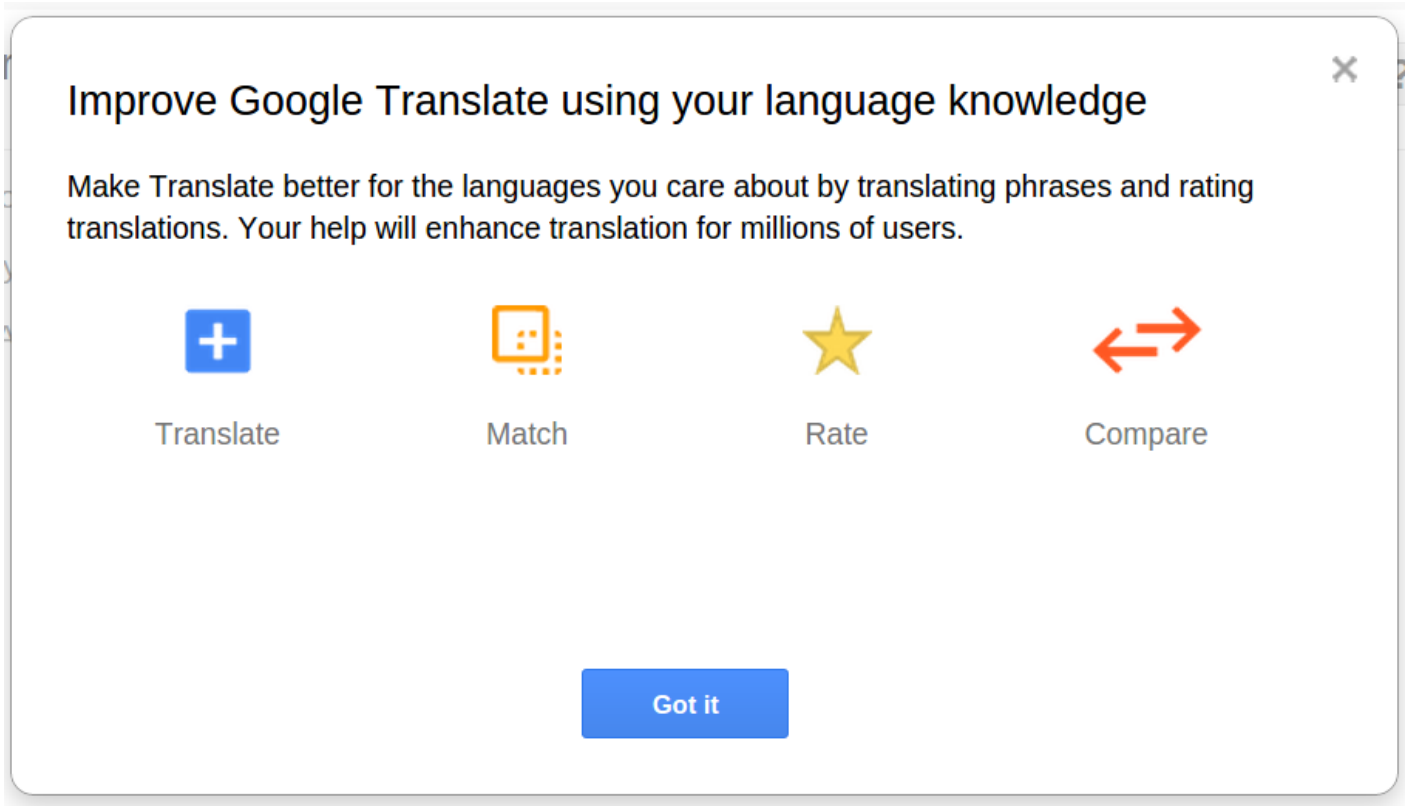
Google Translate: Crowd-Translating

Quindi tutti collaborano alla traduzione, Translate vi propone le traduzioni più frequenti/probabili



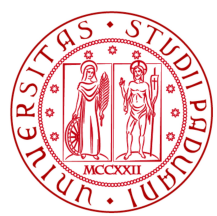
Google Translate: Crowd-Translating

Quindi tutti collaborano alla traduzione, Translate vi propone le traduzioni più frequenti/probabili
Infatti siete i benvenuti:

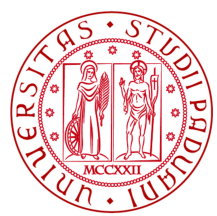


Google Translate: Possibili sviluppi futuri

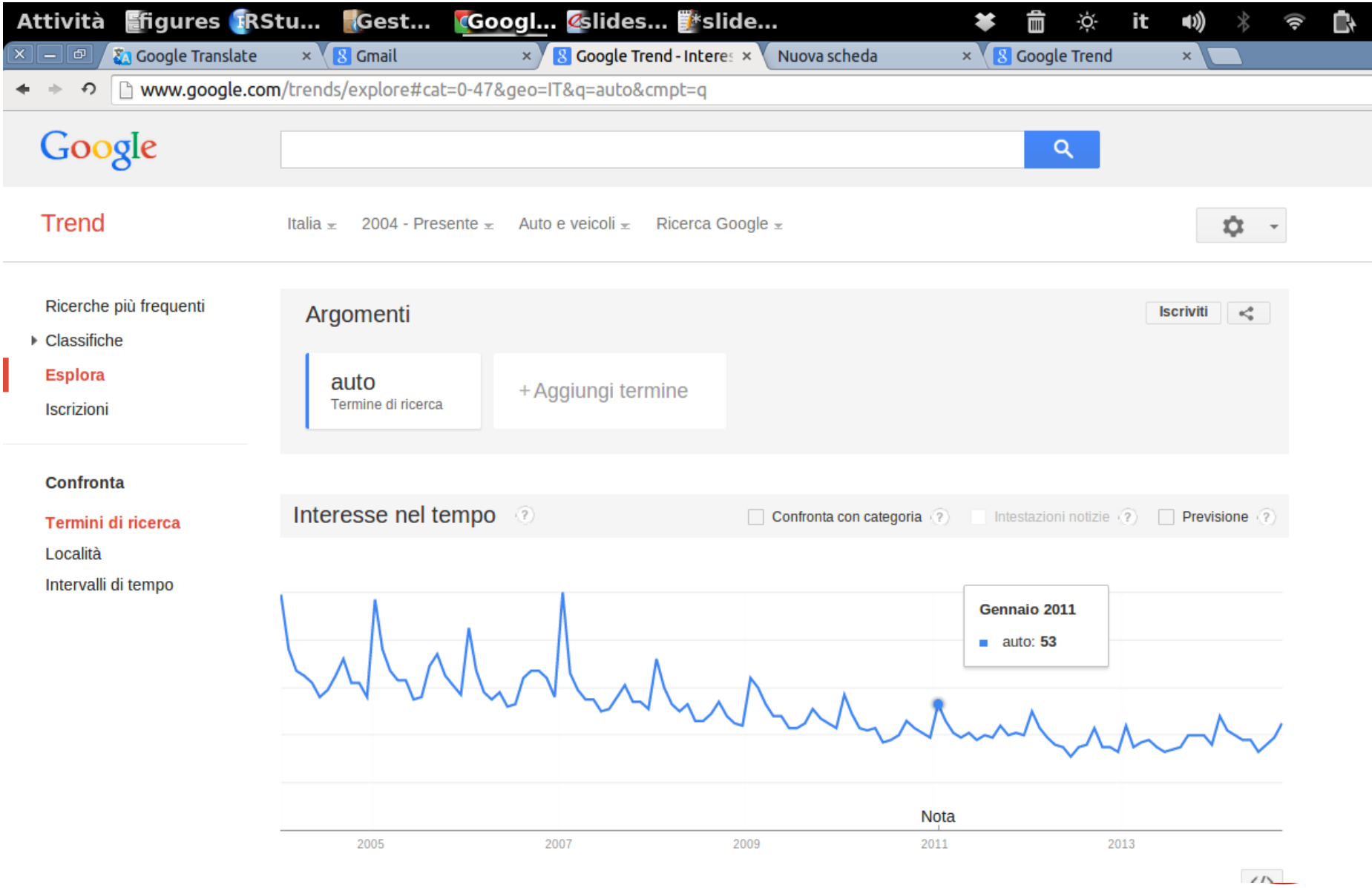
- Il sistema è in grado quindi di auto migliorarsi grazie all'acquisizione di nuovi documenti. Ci possiamo aspettare quindi traduzioni sempre più adeguate.
- traduzioni personalizzate in funzione del vocabolario dell'utente? del tipo di utente?



Google Trends: trends.google.com



Google Trends: trends.google.com

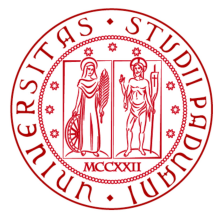


Google Trends: trends.google.com

Per ogni chiave di ricerca (es. 'auto') mostra
il **numero di ricerche**[1] effettuate su google.com dagli utenti (italiani o mondiali).
Ci può aiutare a studiare l'andamento di fenomeni economici e/o sociali.

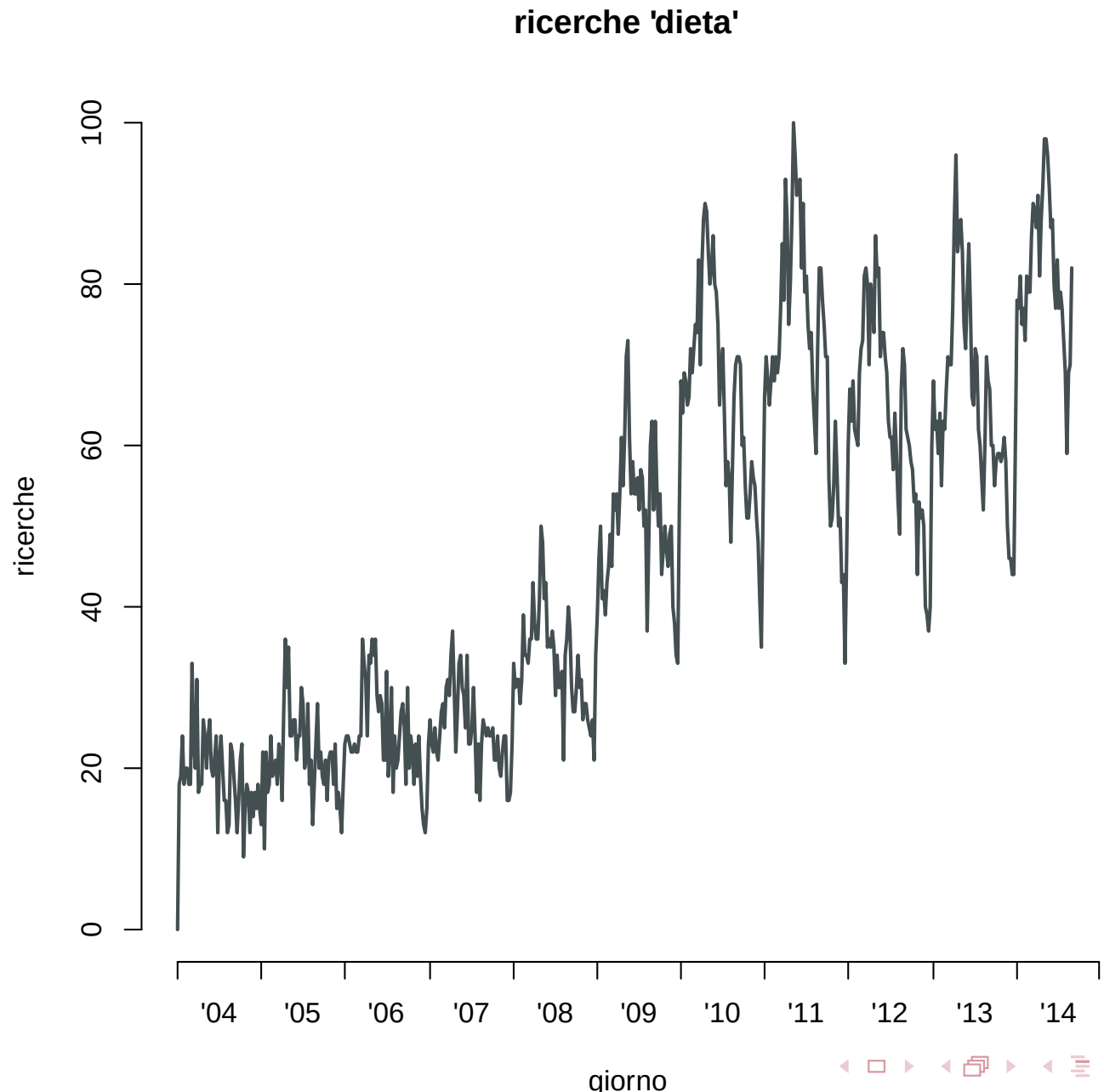
- interesse delle persone per l'acquisto di beni (es. l'auto)
- nascita nuove mode e abitudini

un esempio...



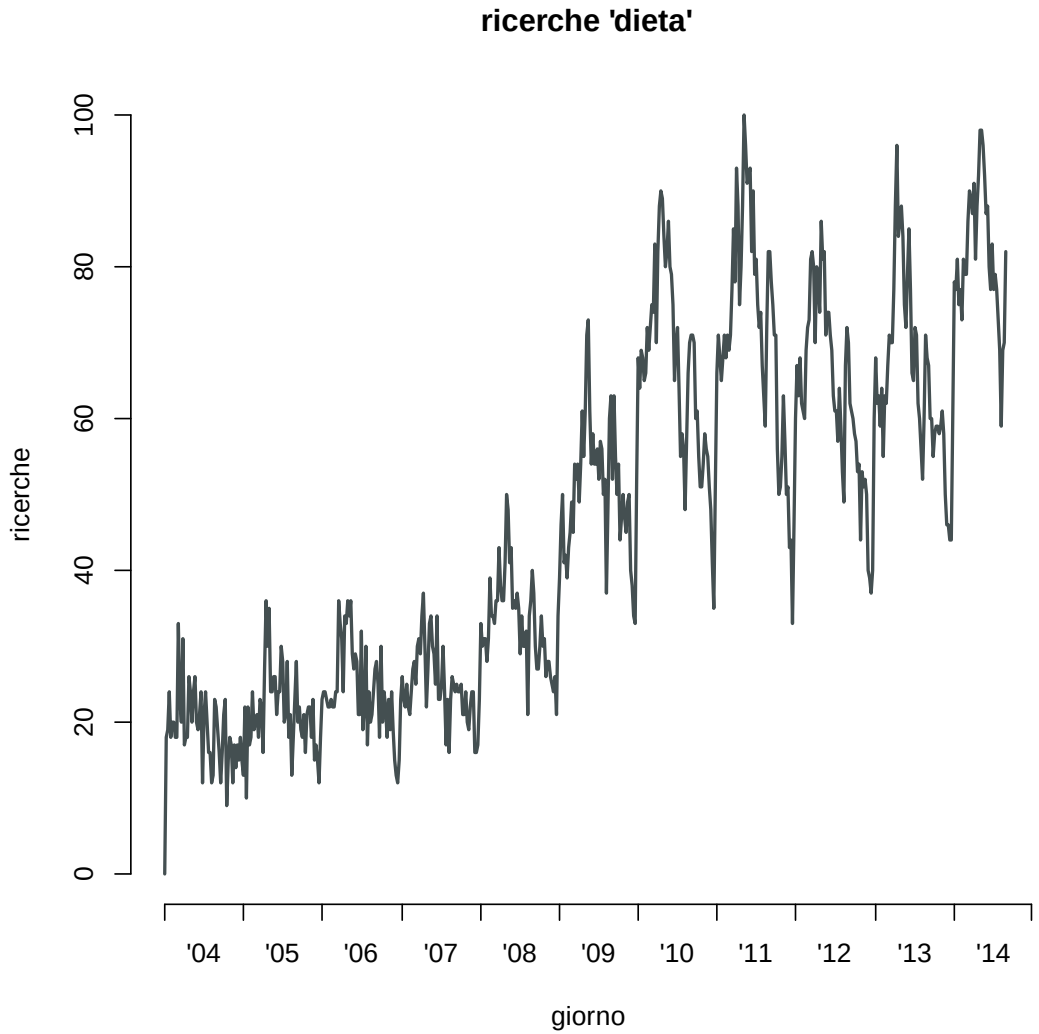
[1] il conteggio in realtà è normalizzato per il numero di ricerche della settimana, si veda trend.google.com per maggiori dettagli

Serie storica delle ricerche della parola 'dieta'



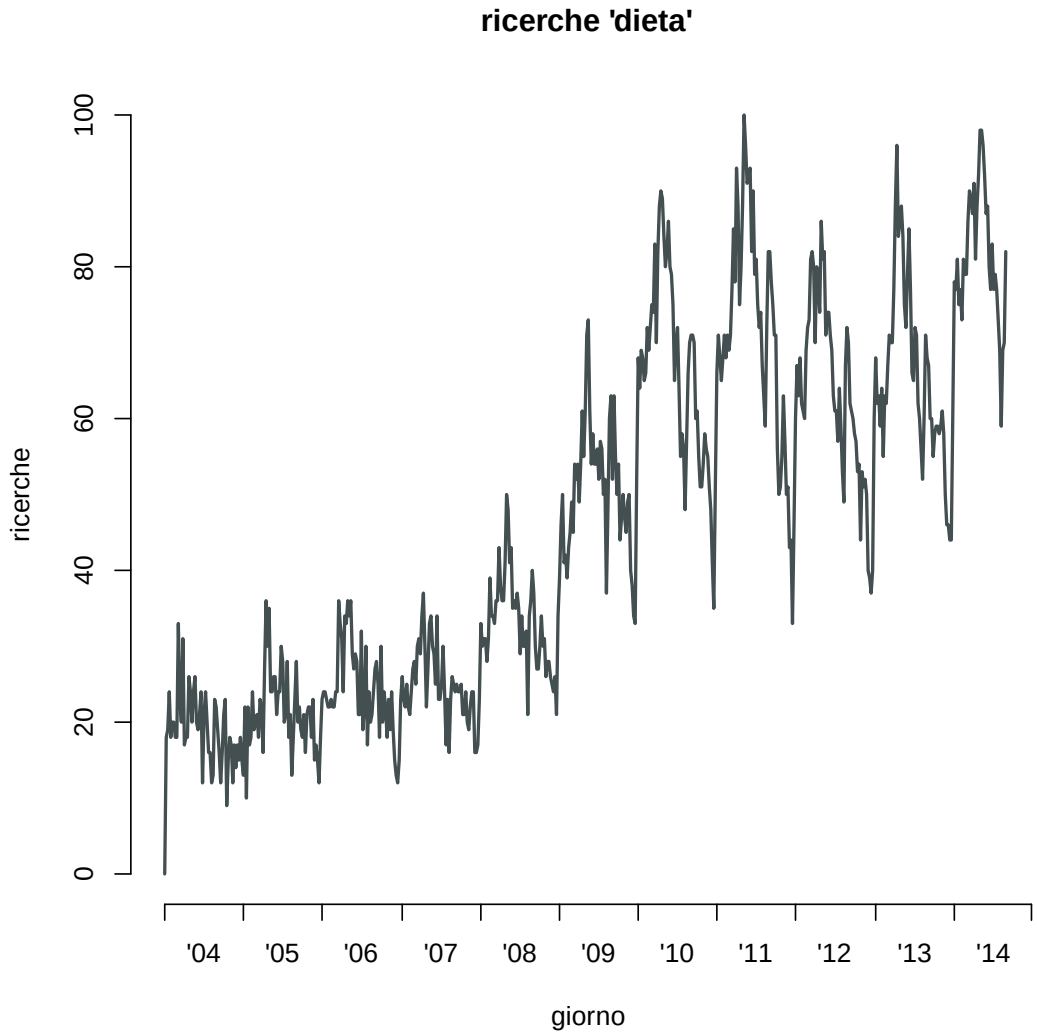
Serie storica delle ricerche della parola 'dieta'

I dati mostrano un 'senso', cerchiamo andamenti 'non casuali'?
(un trend medio, dei pattern tipici etc?)



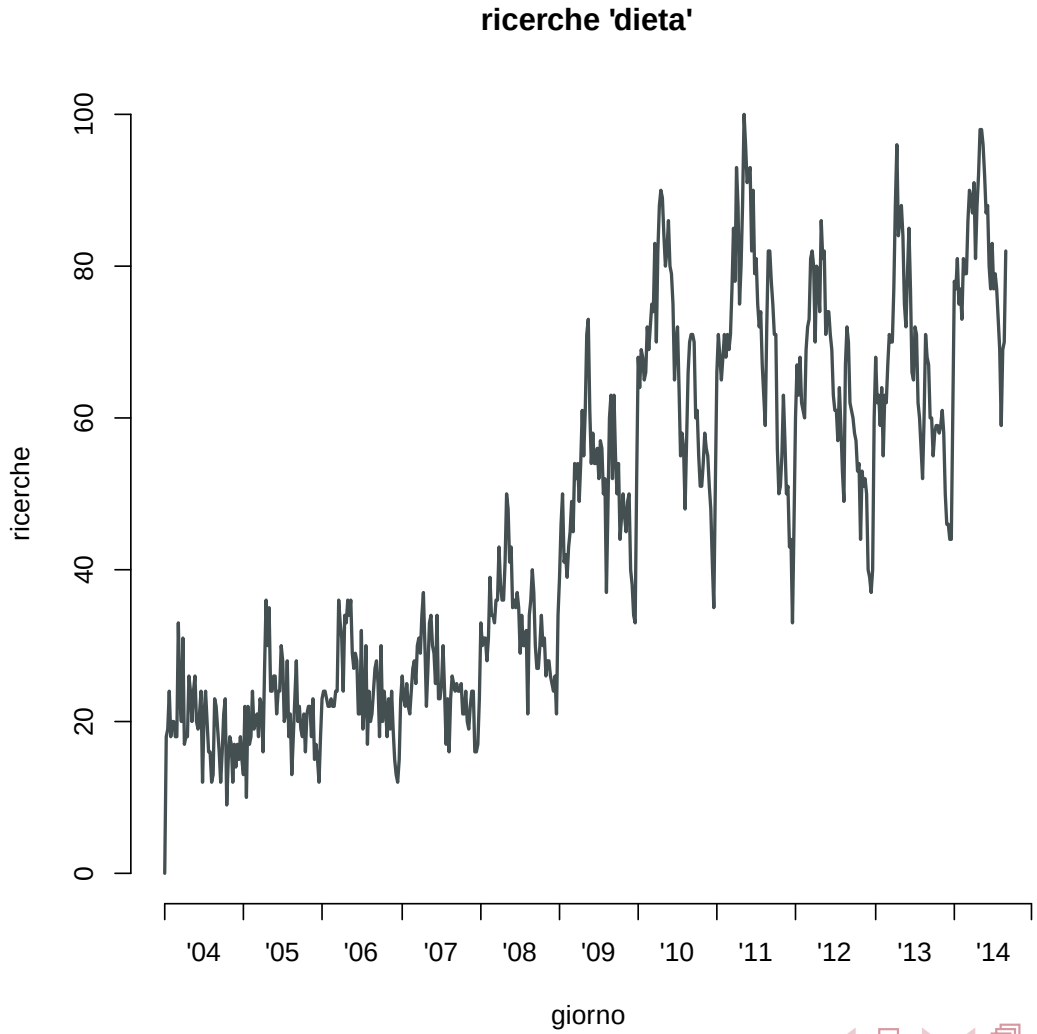
Serie storica delle ricerche della parola 'dieta'

Notiamo **trend** crescente,
cerchiamo di descriverlo con una curva analitica



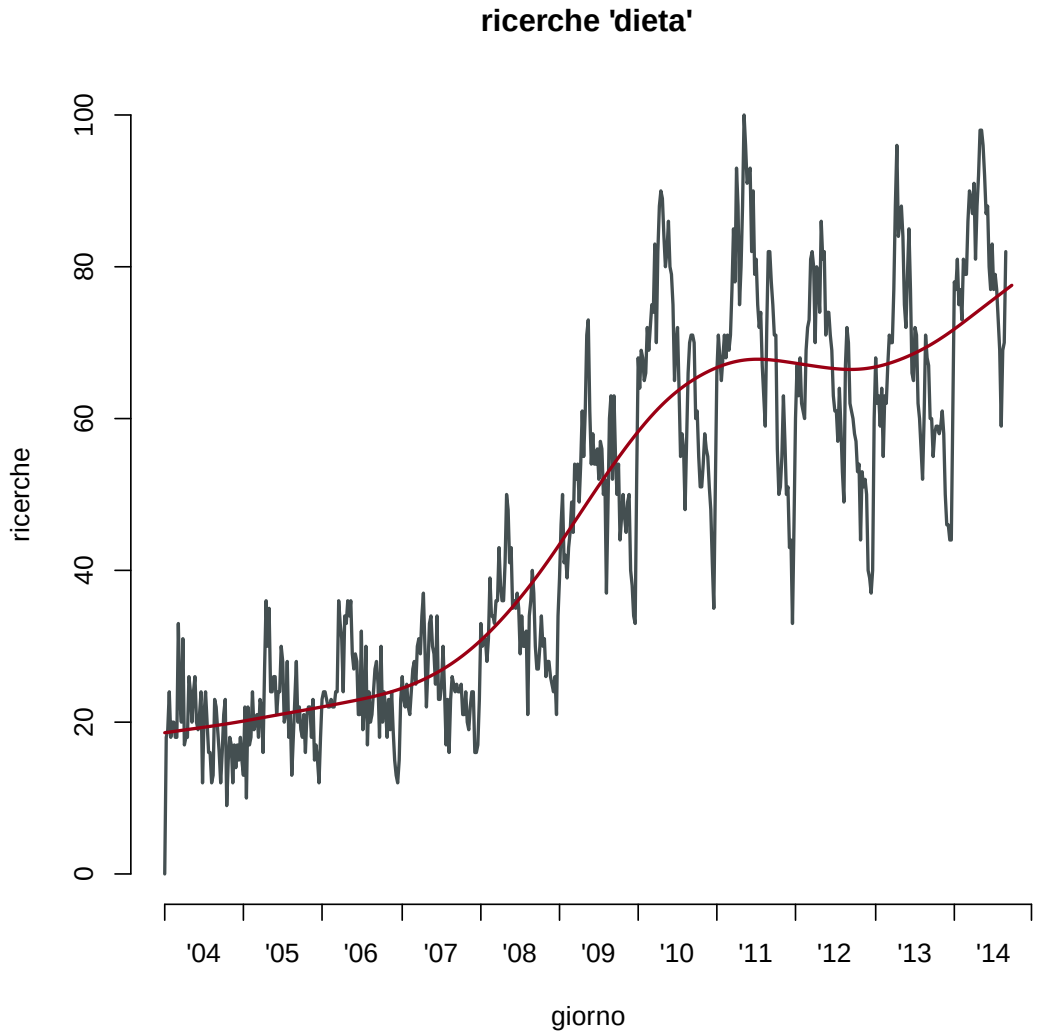
Serie storica delle ricerche della parola 'dieta'

Cerchiamo un modello matematico
$y = f(x)$ cioè: ricercheDimagrire= f(giornoMeseAnno)
che passi (il più possibile) al centro dei dati.



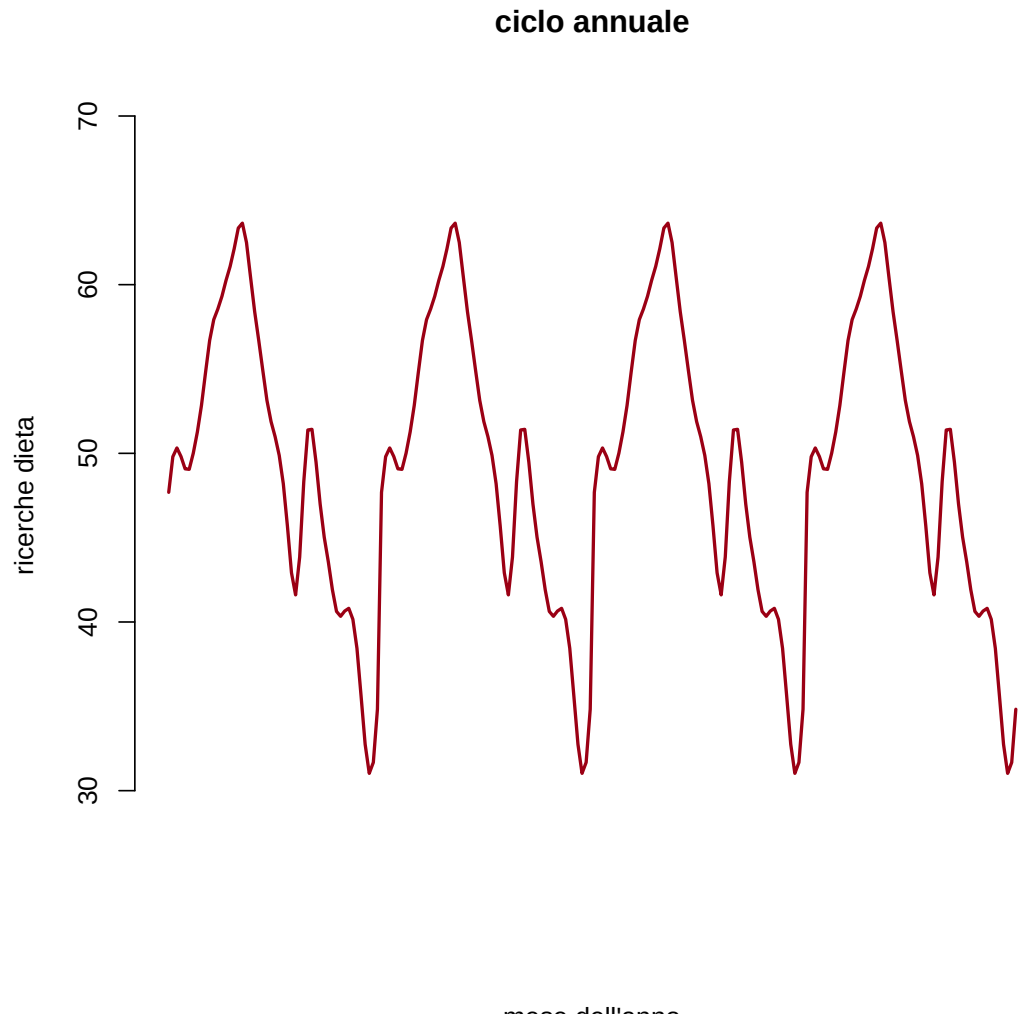
Serie storica delle ricerche della parola 'dieta'

Ci pare che ci sia un andamento che si ripete all'interno dell'anno: **ciclo annuale** non descritto dalla nostra funzione...

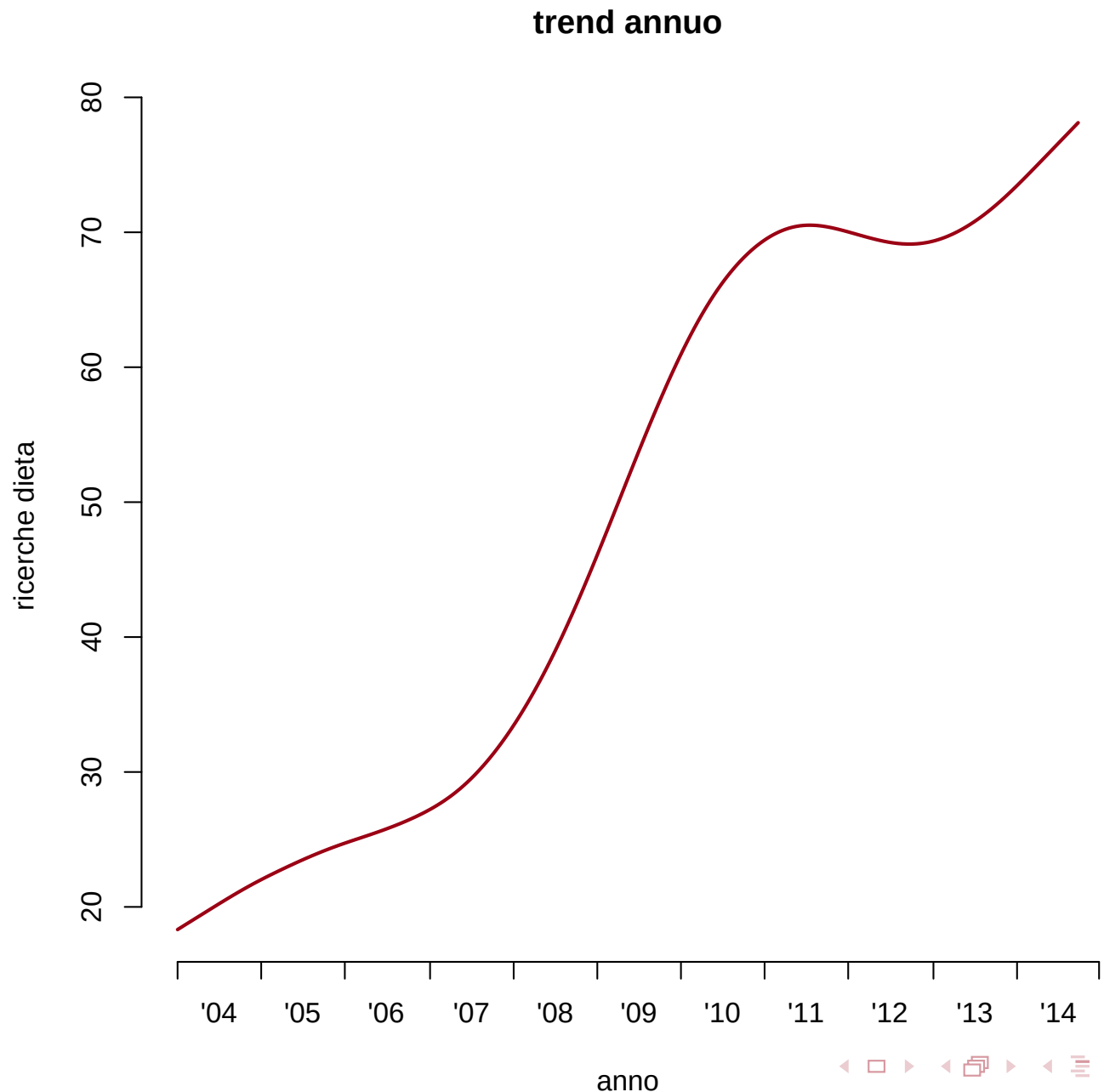


Ciclo annuale

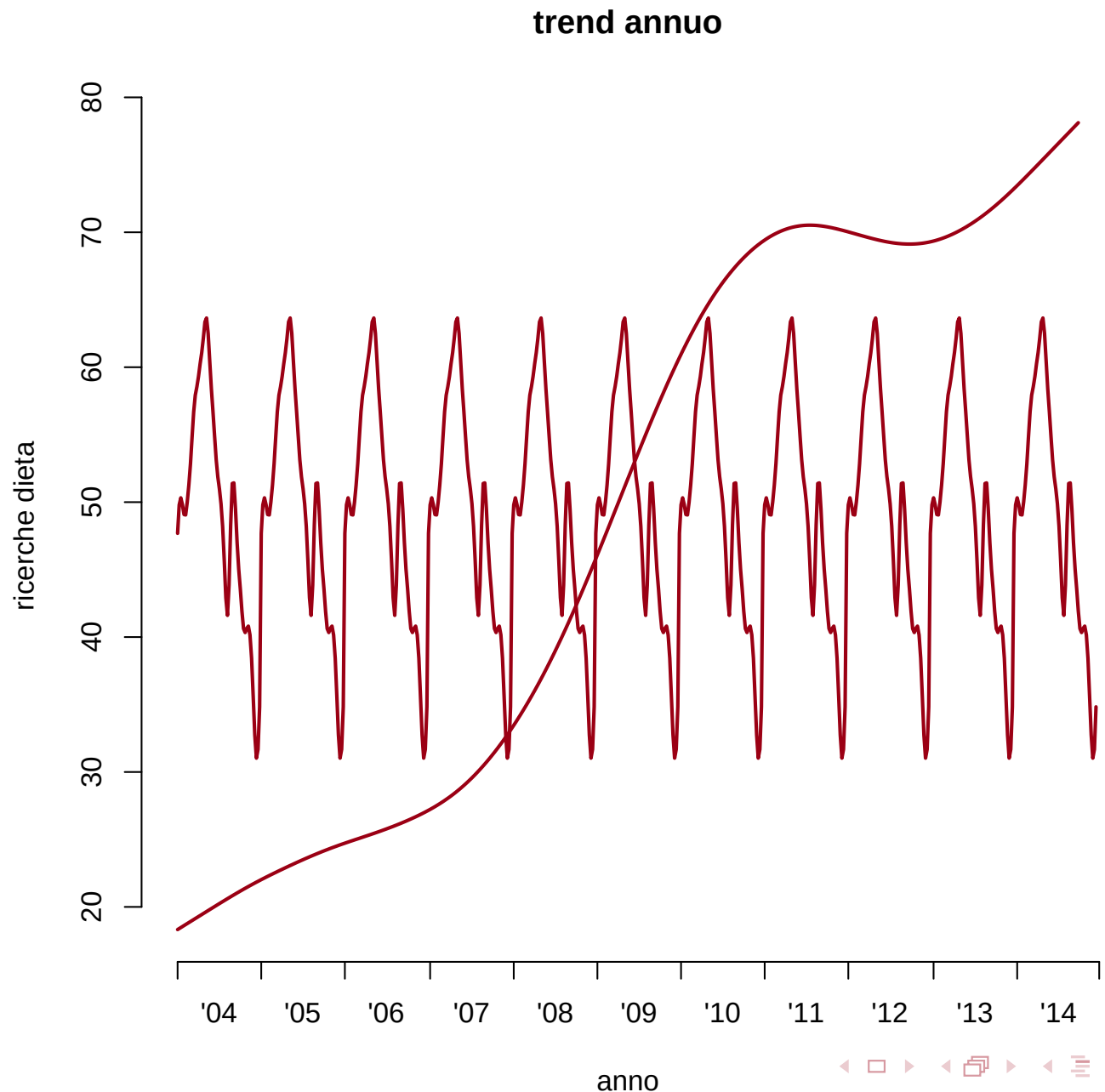
Cerchiamo una funzione $y = g(x)$ periodica (periodo un anno, 365 giorni)
che descriva l'andamento all'interno di ogni anno.



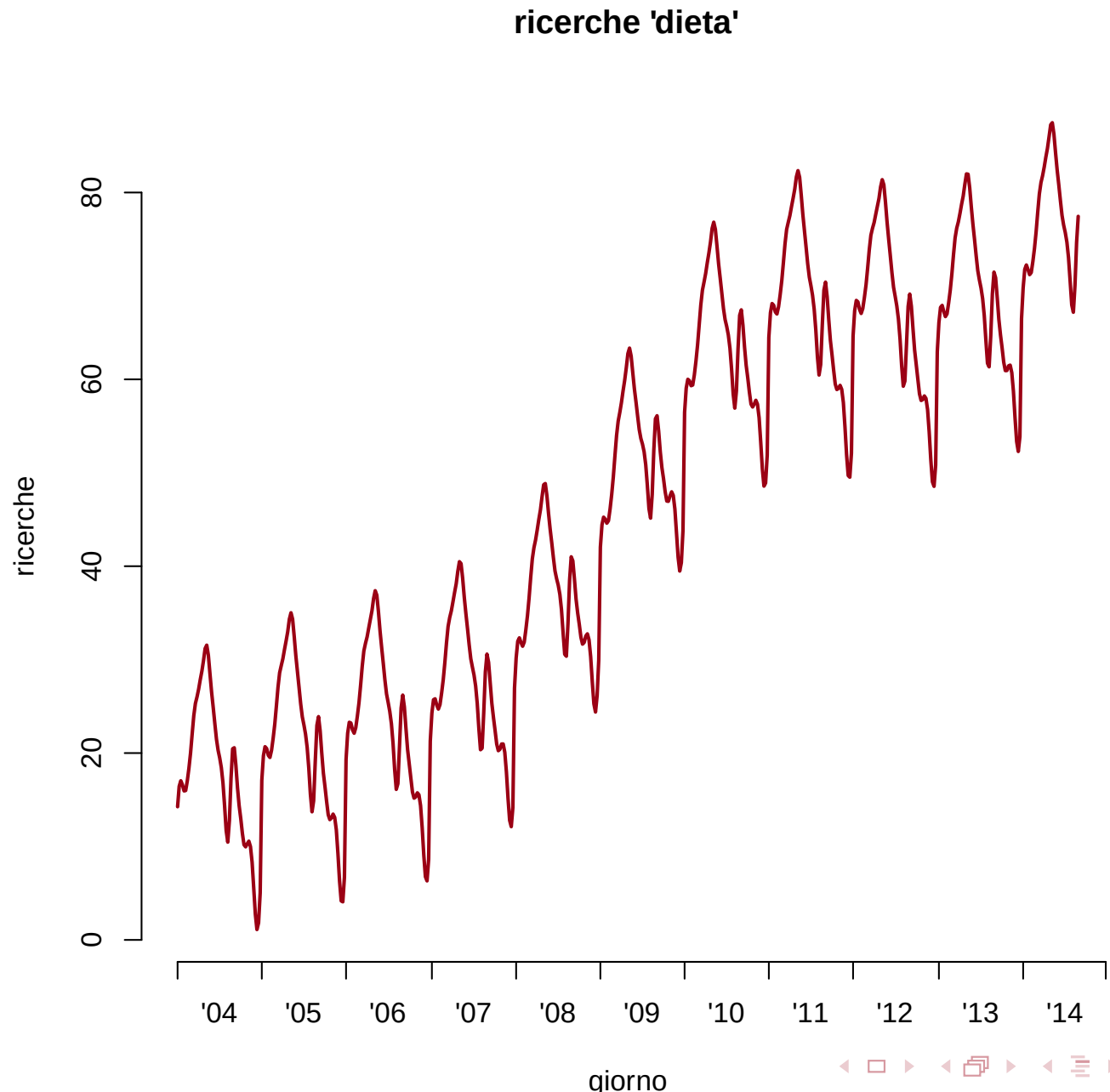
Trend


trend annuo
ricerche dieta
20
30
40
50
60
70
80
'04
'05
'06
'07
'08
'09
'10
'11
'12
'13
'14
anno

f(x) + g(x) = Trend + Ciclo annuale

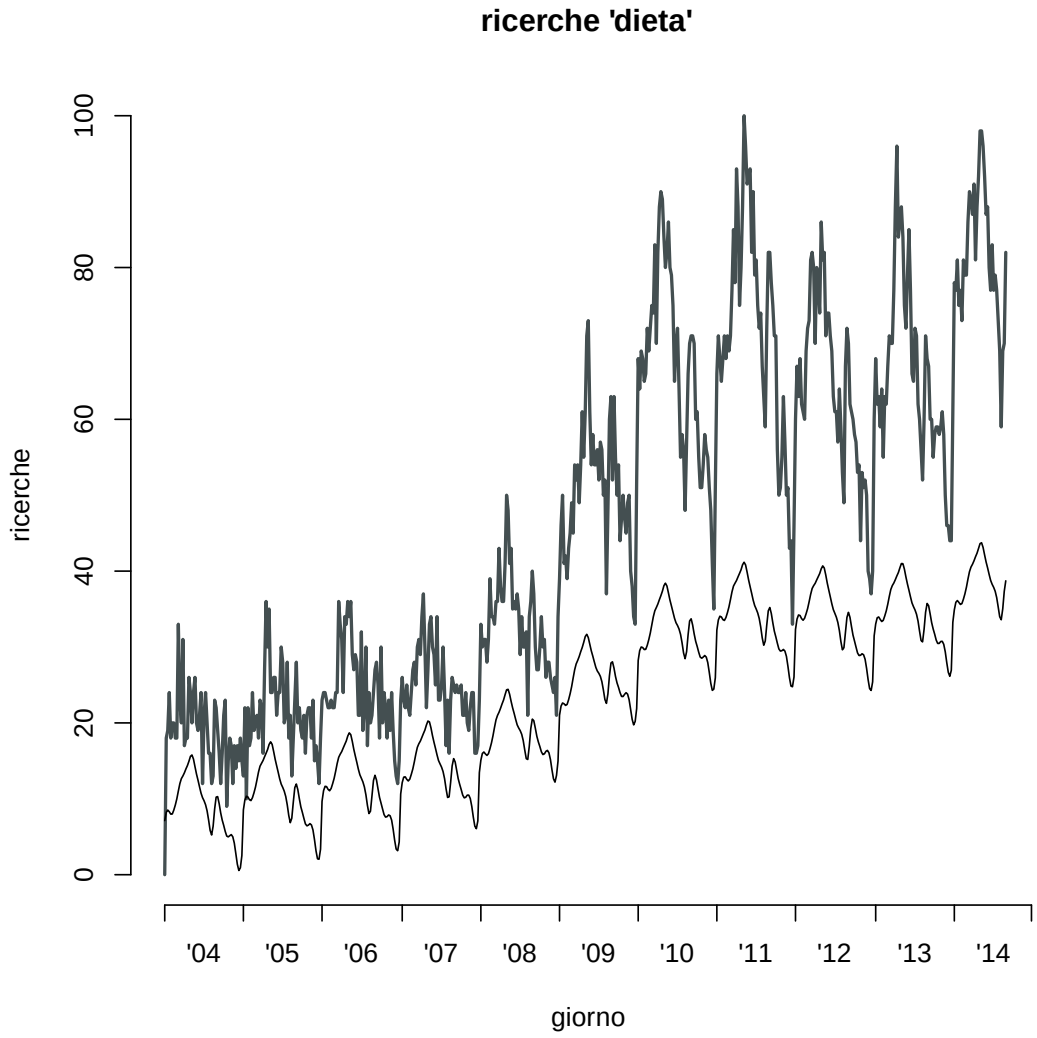


Curva dei valori predetti: $y = f(x) + g(x)$



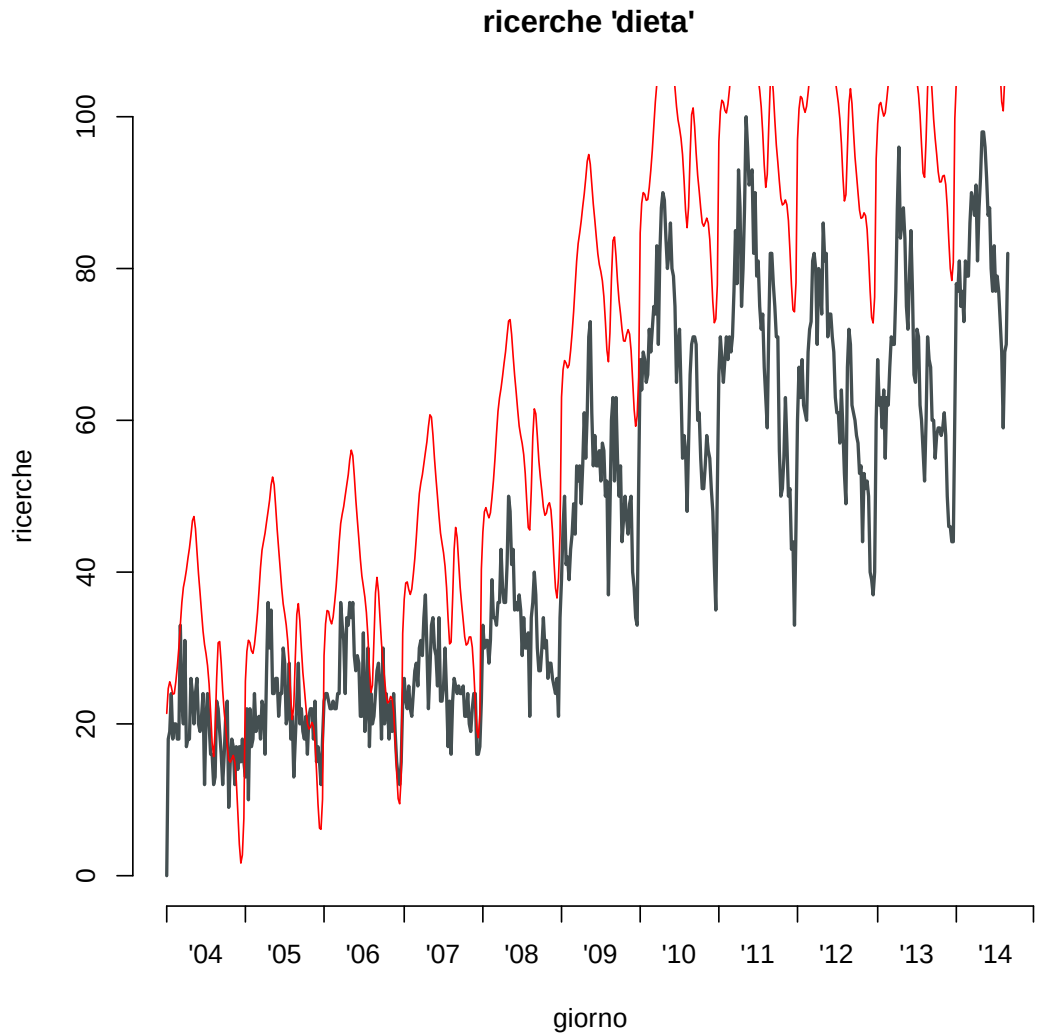
Massima aderenza ai dati

Cerchiamo tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.



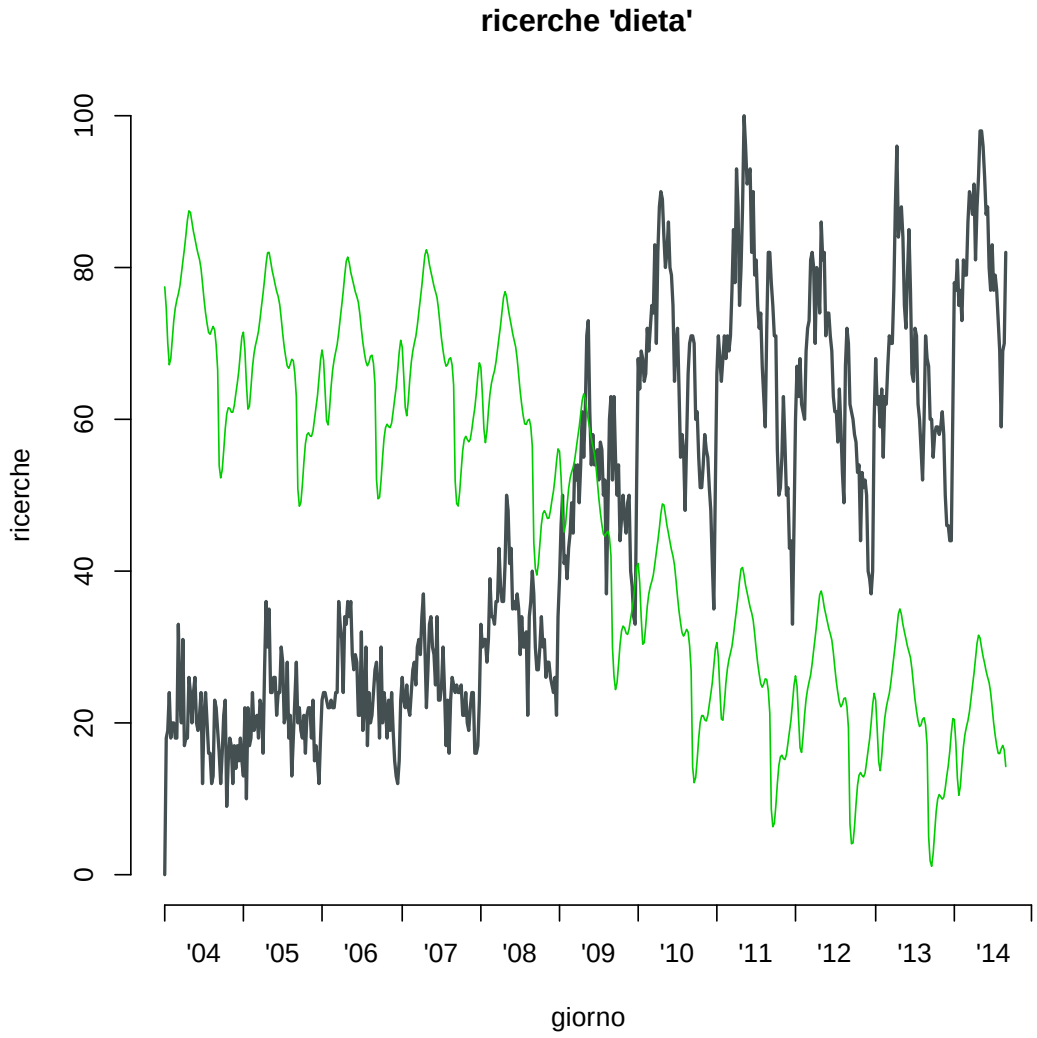
Massima aderenza ai dati

Cerchiamo tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.



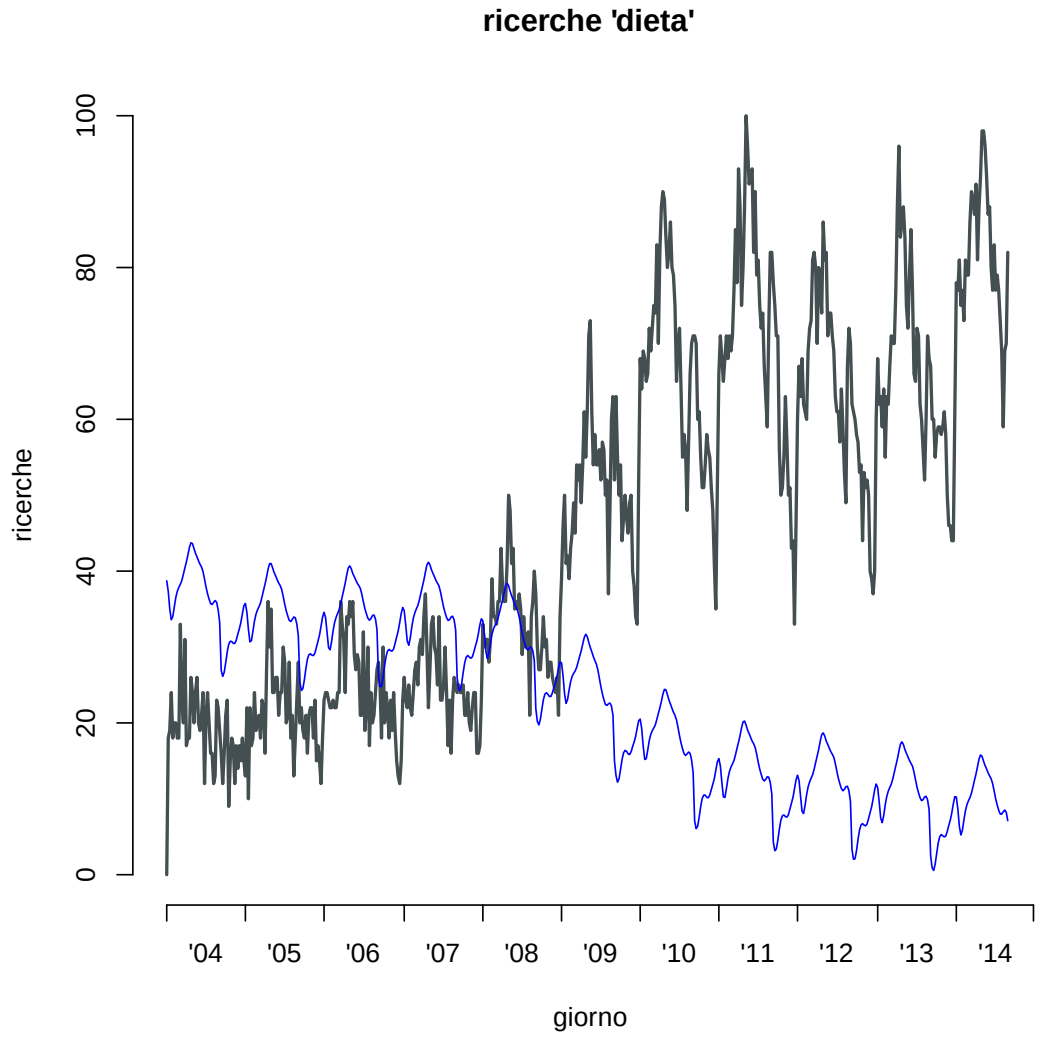
Massima aderenza ai dati

Cerchiamo tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.



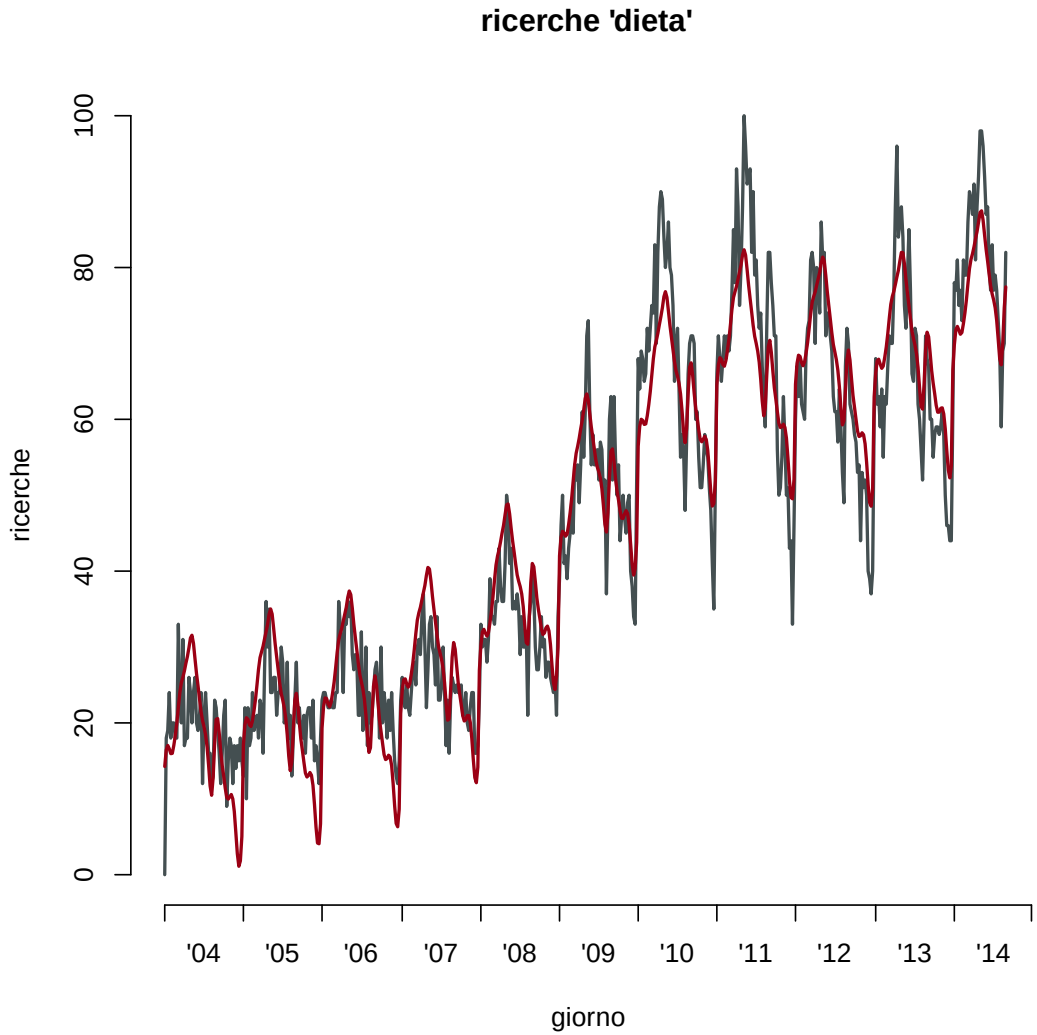
Massima aderenza ai dati

Cerchiamo tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.



Massima aderenza ai dati

Cerchiamo tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.

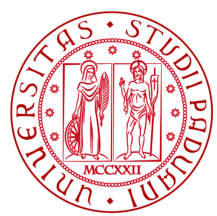


Massima aderenza ai dati: Metodo dei Minimi Quadrati

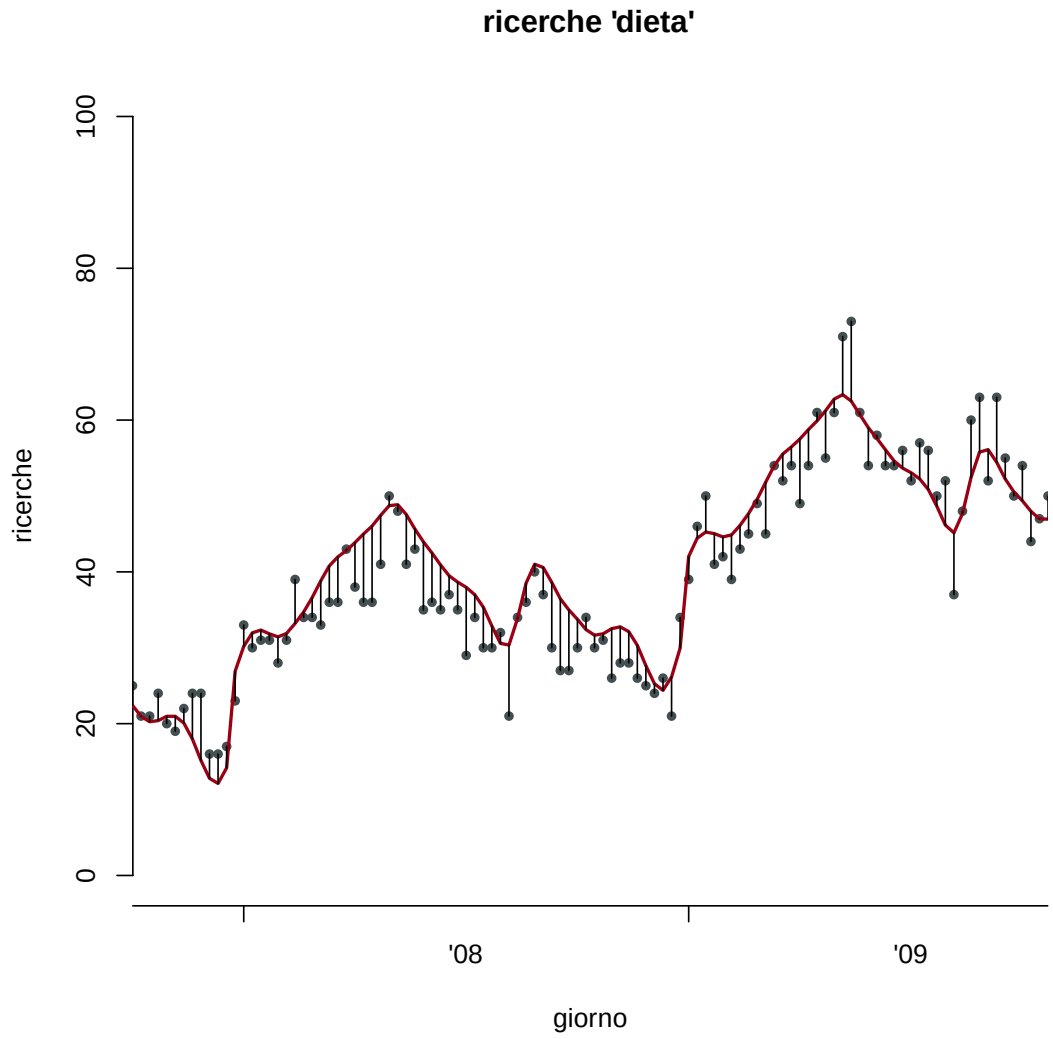
Abbiamo cercato tra tutte le curve possibili, la coppia (trend e ciclo) che passa meglio intorno ai dati.
In termini (un po') più formali:

- numero ricerche: $y_1, \ldots, y_N$
- giorno (la data): $x_1, \ldots, x_N$
- numero ricerche predette dal modello: $\hat{y}_i = f(x_i) + g(x_i)$ (abbiamo quindi $\hat{y}_1, \ldots, \hat{y}_N$)
  - cerchiamo le funzioni f e g tali che $\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ sia minima.

In alcuni casi la soluzione è analitica (derivate e punti di minimo), altre volte è numerica (algoritmi e computer).

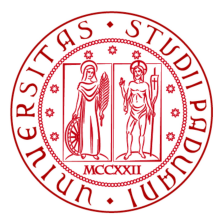


Massima aderenza ai dati: Metodo dei Minimi Quadrati



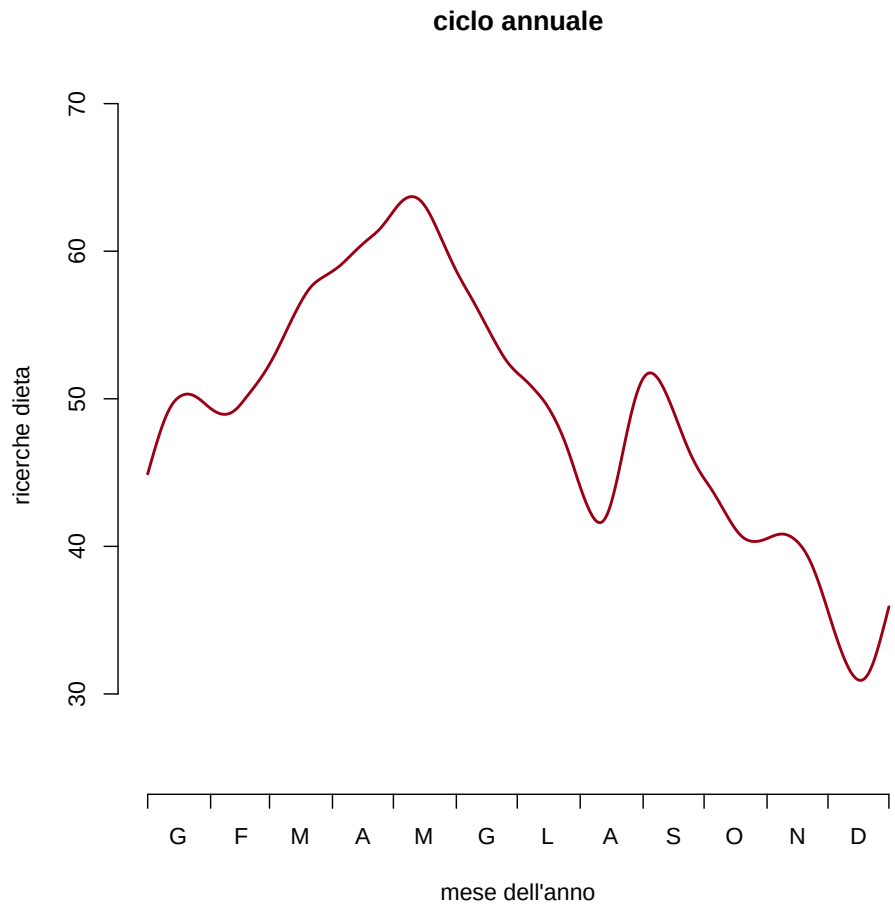
Interpretazione 'sociologica' del fenomeno

Il modello identifica **due fenomeni** che determinano il numero di ricerche della parola 'dieta':



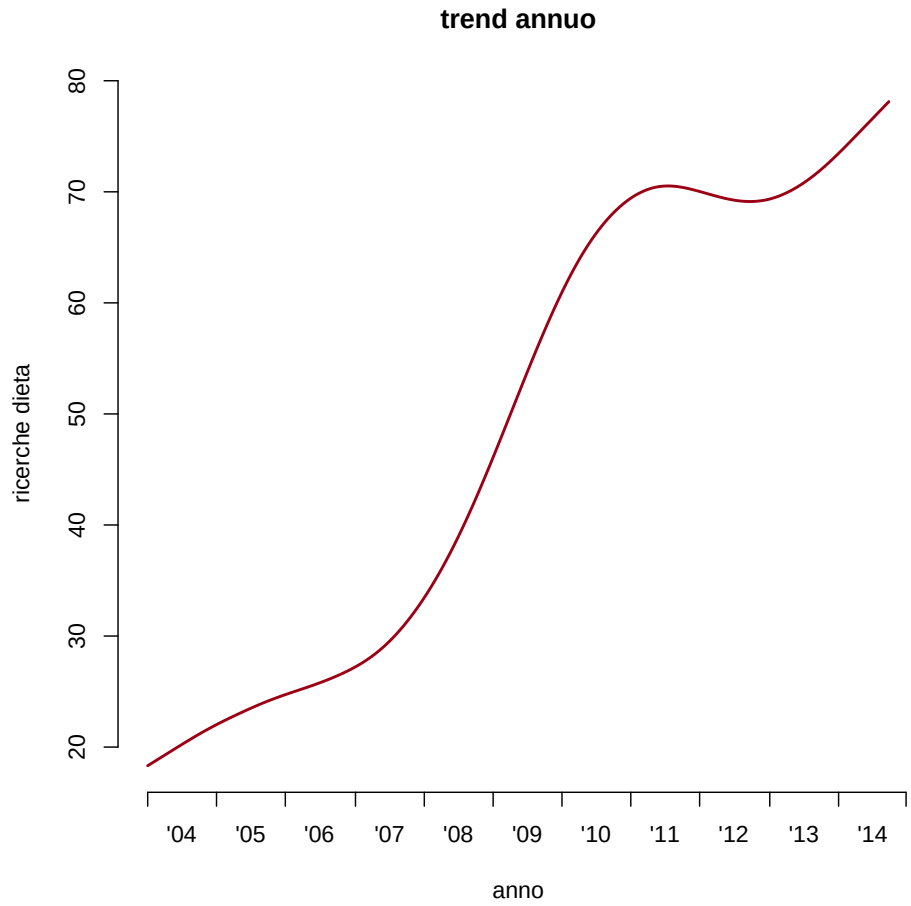
Interpretazione 'sociologica' del fenomeno

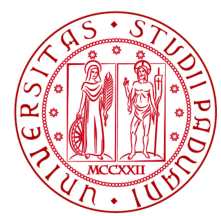
1) **Ciclo annuale**: non ci si mette in dieta a Natale, lo si fa a maggio e ci si riprova a settembre con meno convinzione)



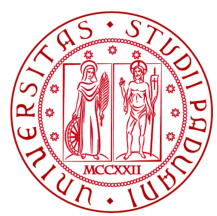
Interpretazione 'sociologica' del fenomeno

2) **Trend**: monotono non decrescente che conferma le preoccupazioni di psicologi e sociologi...





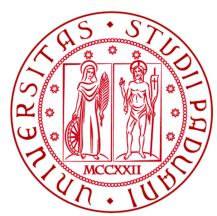
Google Trends: Previsione di fenomeni reali



Google Trends: Previsione di fenomeni reali

Ci può aiutare anche a studiare l'andamento di fenomeni economici e/o sociali reali e misurabili.

- disoccupazione (più persone cercano lavoro, maggiore è il numero di ricerche su google)
- l'insorgenza e la diffusione dell'influenza (...)



Google Flu

Home page di Google.org
(inglese)

Diffusione della dengue

Trend influenzali

Home page

Seleziona Paese/regione ▾

Come funziona?

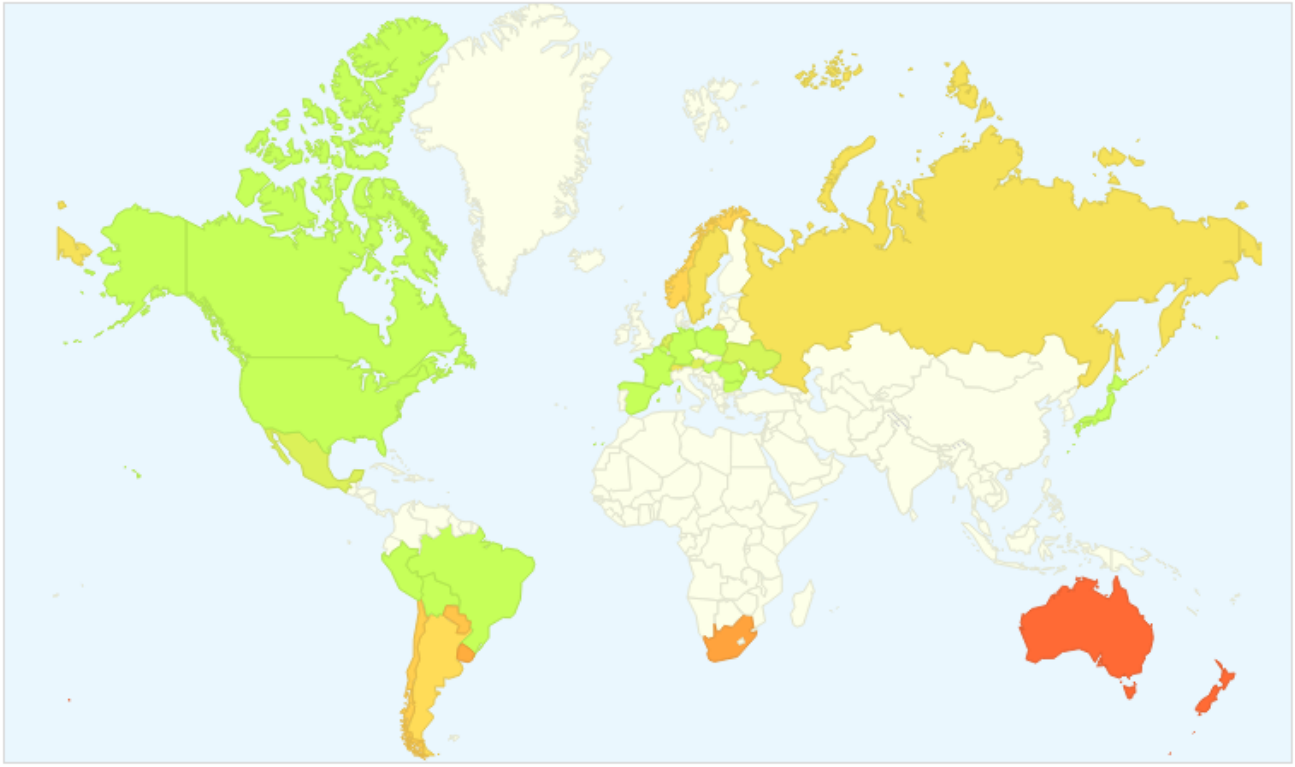
Domande frequenti (FAQ)

Attività influenzale

- Intensa
- Alta
- Moderata
- Bassa
- Minima

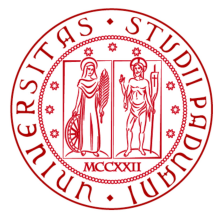
Scopri i trend influenzali nel mondo

Abbiamo scoperto che determinati termini di ricerca sono validi indicatori dell'attività influenzale. Google Trend influenzali utilizza dati di ricerca aggregati di Google per stimare l'attività influenzale. Ulteriori informazioni »



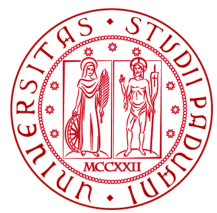
Google Flu: come funziona

- Seleziona le ricerche che aumentano quando si sta diffondendo l'influenza (es. 'sintomi influenzali', 'curare l'influenza', 'quanto dura l'influenza')
  abbiamo quindi una serie storica di ricerche su google

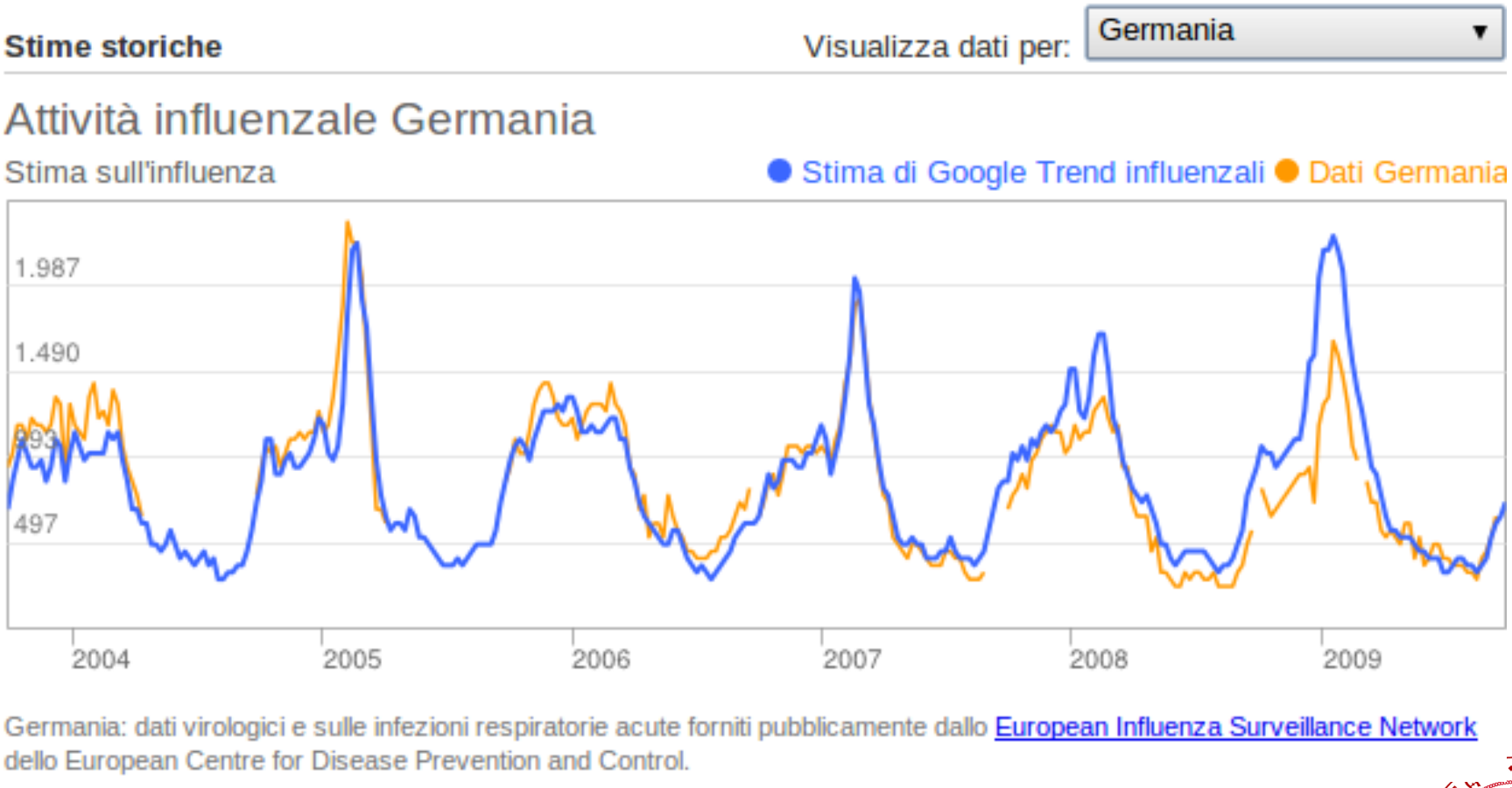


Google Flu: come funziona

- Seleziona le ricerche che aumentano quando si sta diffondendo l'influenza (es. 'sintomi influenzali', 'curare l'influenza', 'quanto dura l'influenza')
  abbiamo quindi una serie storica di ricerche su google
- riscala la serie storica in modo che stia più vicina possibile alla serie storica del numero di influenzati:
  - numero di influenzati ogni giorno: $y_1, \ldots, y_N$
  - numero di ricerche su google: $x_1, \ldots, x_N$
  - dati predetti dal modello (ad es.) $\hat{y}_i = f(x_i) = a + b\, x_i$
  - cerchiamo la coppia di valori a e b tali per cui $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$ sia minima. (soluzione analitica!)

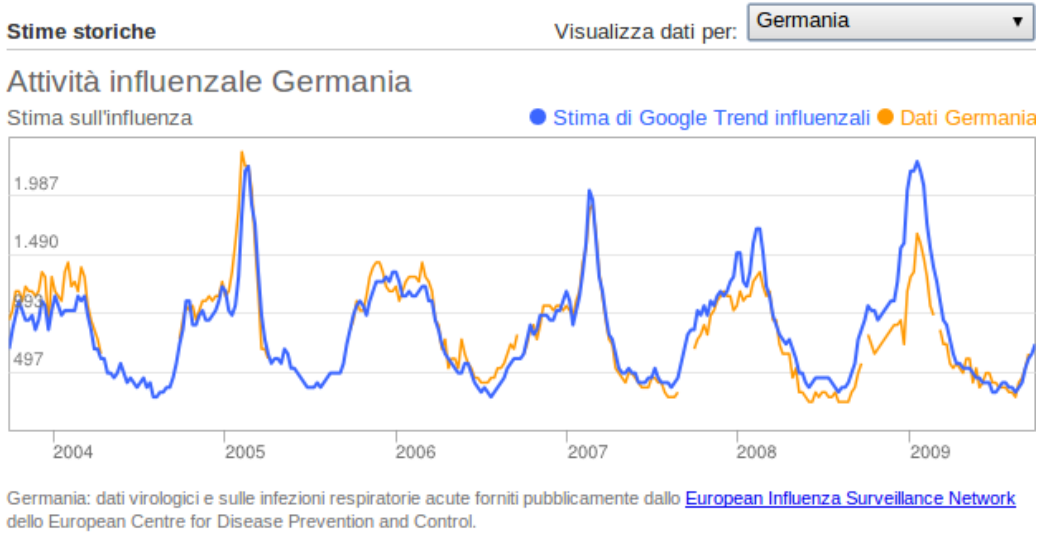


Valori predetti vs reali



Similitudini e diversità da spiegare..

Germania



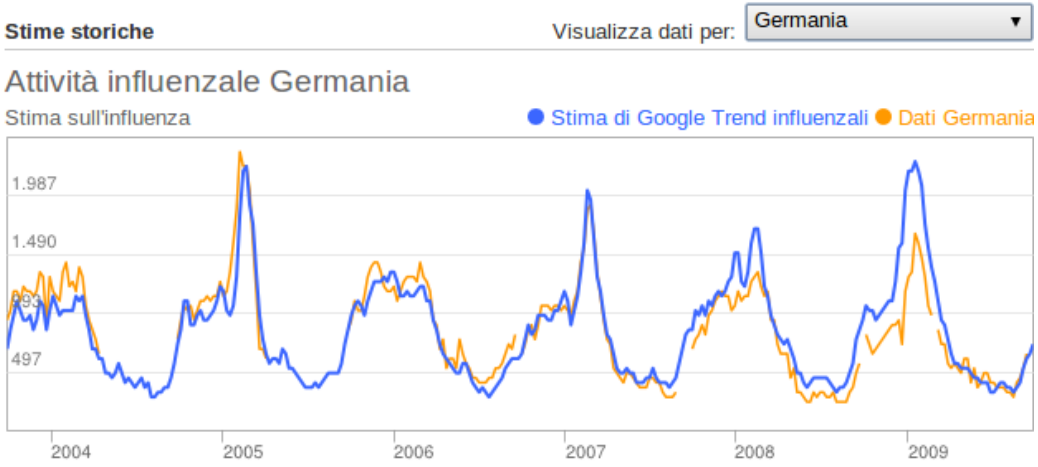
Germania: dati virologici e sulle infezioni respiratorie acute forniti pubblicamente dallo European Influenza Surveillance Network dello European Centre for Disease Prevention and Control.

vs Austria

Austria: dati ILI (Influenza-Like Illness) forniti pubblicamente dallo European Influenza Surveillance Network dello European Centre for Disease Prevention and Control.

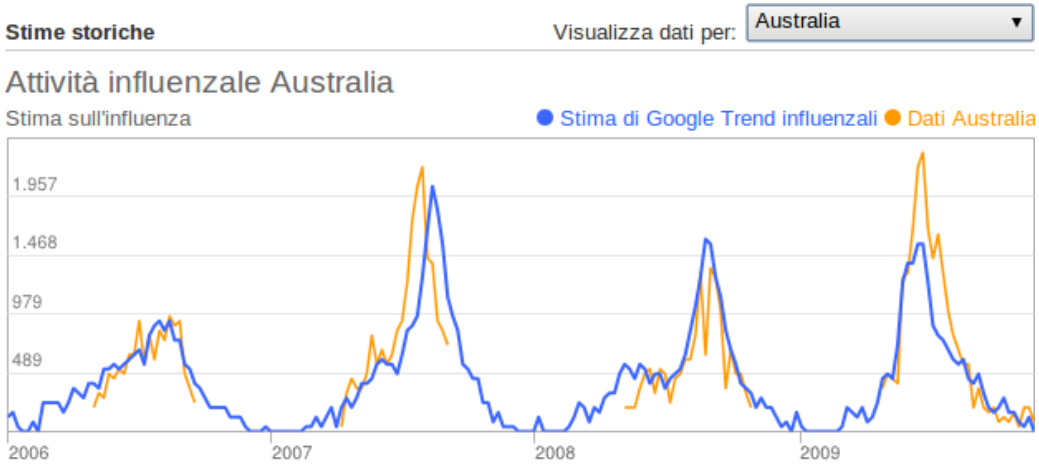
Similitudini e diversità da spiegare..

Germania



Germania: dati virologici e sulle infezioni respiratorie acute forniti pubblicamente dallo European Influenza Surveillance Network dello European Centre for Disease Prevention and Control.

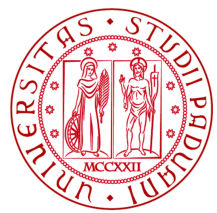
vs Australia



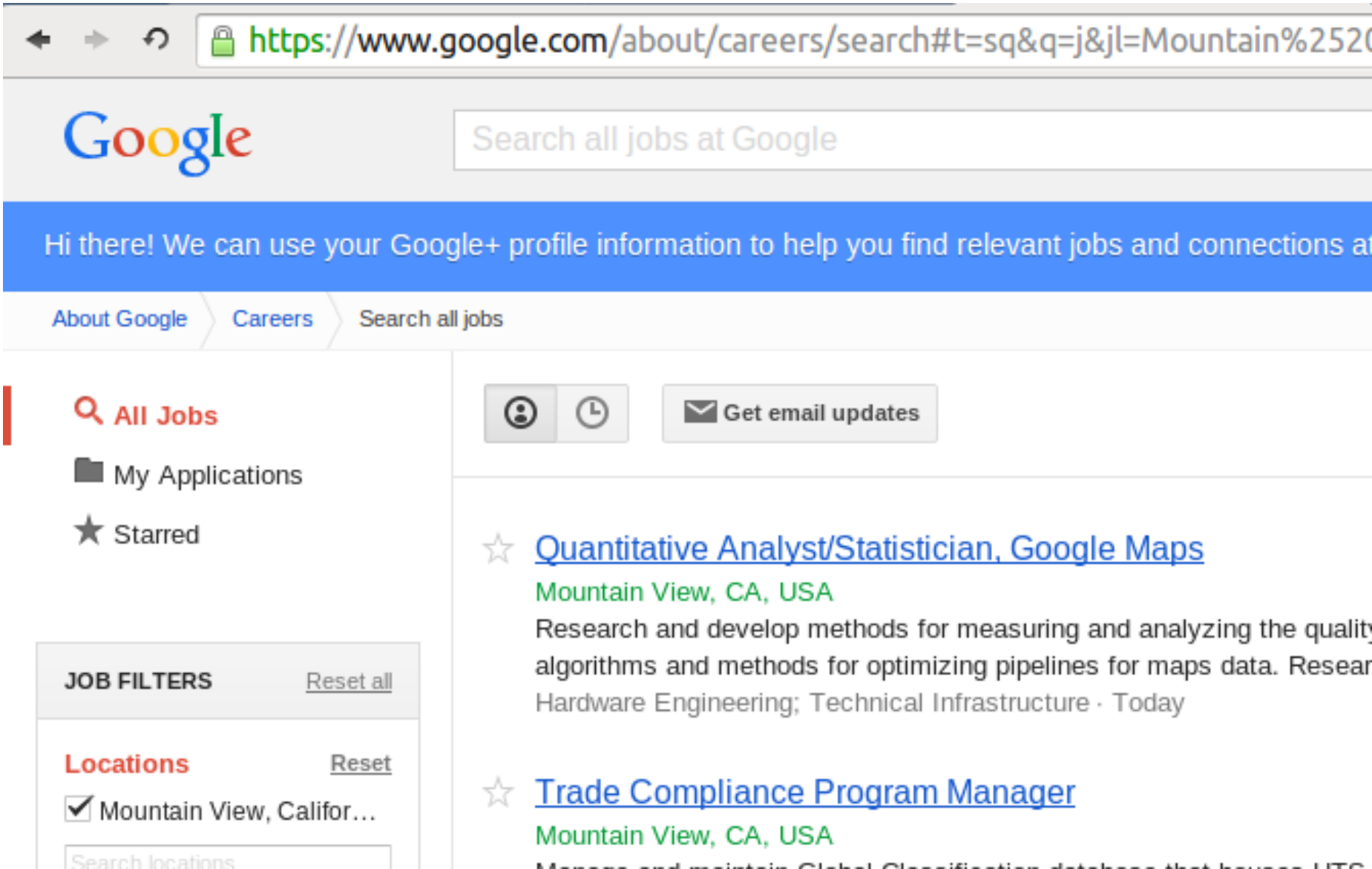
Australia: dati ILI (Influenza-Like Illness) forniti pubblicamente dal Victorian Infectious Diseases Reference Laboratory.

Il potere del modello

una volta che abbiamo definito la funzione (= il modello)
$\hat{y}_i = f(x_i) = a + b\ x_i$
es. #Influenze= $1100 + 100$ #Ricerche
possiamo fare previsioni (accurate?) e in tempo reale del numero di influenzati (solitamente i servizi sanitari ci mettono molto tempo a pubblicare i dati).



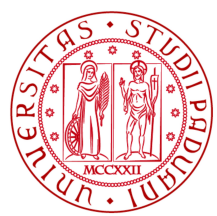
ben si capisce perché Google cerca statistici/data scientist



Sentiment Analysis

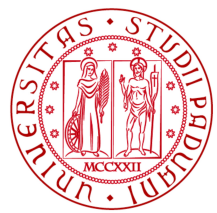
Sentiment Analysis

- Quando un post (ad es di politica, di musica, di prodotti commerciali) riceve migliaia di commenti diventa molto difficile leggerli e sintetizzarli per capire cosa ne pensano gli utenti.



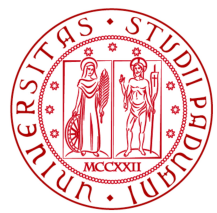
Sentiment Analysis

- Quando un post (ad es di politica, di musica, di prodotti commerciali) riceve migliaia di commenti diventa molto difficile leggerli e sintetizzarli per capire cosa ne pensano gli utenti.
- Eppure sarebbe molto utile per orientare le scelte politiche, commerciali etc.

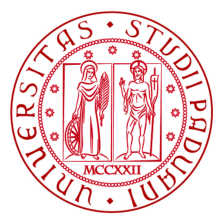


Sentiment Analysis

- Quando un post (ad es di politica, di musica, di prodotti commerciali) riceve migliaia di commenti diventa molto difficile leggerli e sintetizzarli per capire cosa ne pensano gli utenti.
- Eppure sarebbe molto utile per orientare le scelte politiche, commerciali etc.
- La sentiment analysis ci può dare una mano.



Sentiment Analysis: Il metodo più semplice



Sentiment Analysis: Il metodo più semplice

Creo un vocabolario di parole **positive** e uno di **negative**

Per ogni commento: + # positive - # negative

se il risultato è un numero positivo/negativo, il commento è considerato positivo/negativo.

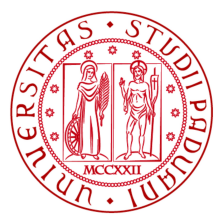
ad es. 1000 commenti: 700 classificati positivi, 300 negativi:
% positivi= $\frac{700}{1000} \times 100 = 70\%$

Problemi:

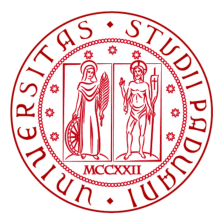
I vocabolari/registri cambiano in funzione dei contesti.
es. la parola **fuoco**

- es musicale: 'questa canzone è fuoco per me!!' (positiva)
- es politico: 'questo è solo un fuoco di paglia' (negativo)

Diventa molto difficile fare analisi efficaci.

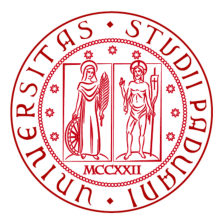


Sentiment Analysis: Il metodo statistico



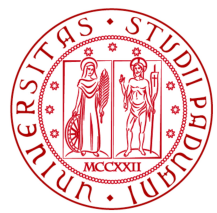
Sentiment Analysis: Il metodo statistico

- **Classifico 'a mano'** qualche centinaia di commenti sullo **specifico argomento** che stiamo analizzando
- (sulla base dei dati sopra) per ogni configurazione di parole **stimo le probabilità** che la frase abbia senso positivo e negativo:
  se $P(\textit{positivo}|\text{frase}) > P(\textit{negativo}|\text{frase})$
  classifico il commento come positivo
- classifico tutti gli **altri commenti** (non classificati 'a mano') e stimo la proporzione totale di commenti positivi.



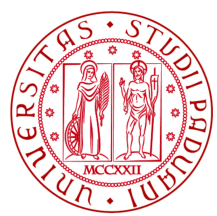
Sentiment Analysis: alcune riflessioni

- Naturalmente il metodo non è perfetto, ma ci permette di classificare moli di dati che non sarebbe possibile valutare 'a mano'.
- Spesso volte è già molto importante avere valori indicativi (es: più positivi che negativi? di molto/poco?)



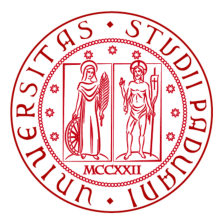
Sentiment Analysis

Possiamo quindi sapere cosa ne pensa (un campione del)la popolazione a proposito



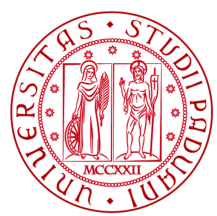
Sentiment Analysis

Possiamo quindi sapere cosa ne pensa (un campione del)la popolazione a proposito
degli immigrati



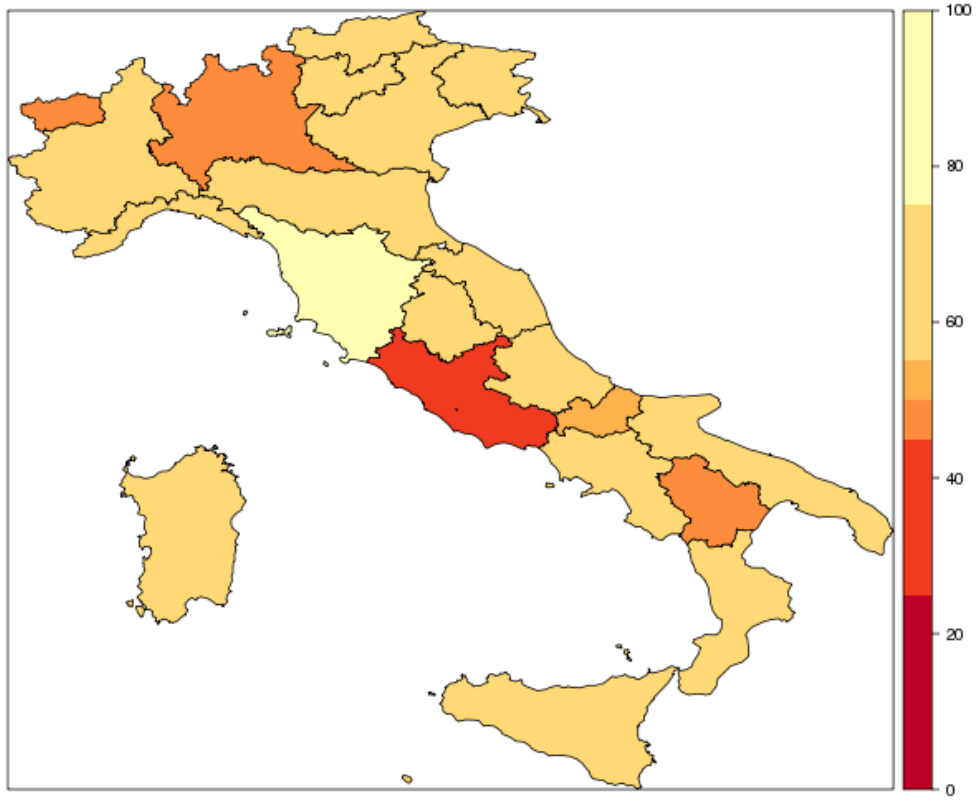
Sentiment Analysis

Possiamo quindi sapere cosa ne pensa (un campione del)la popolazione a proposito
degli immigrati o dei Dear Jack

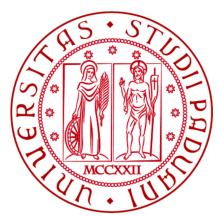


Sentiment Analysis

Possiamo quindi sapere cosa ne pensa (un campione del)la popolazione a proposito
degli immigrati o dei Dear Jack
oppure quali sono le regioni più felici
(`voicesfromtheblogs.com`, dati `Twitter.com`)



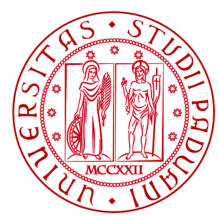
Dalla Sentiment Analysis alla Content Analysis



Dalla Sentiment Analysis alla Content Analysis

Si possono anche avere informazioni simili a quelle delle indagini campionarie classiche

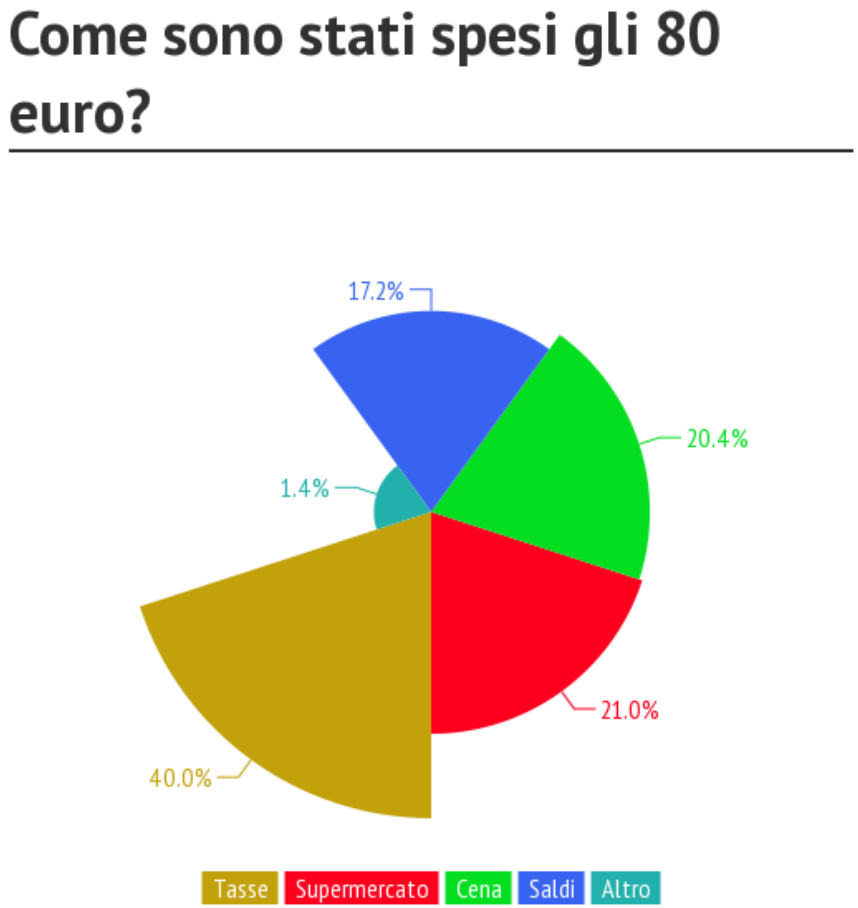
ad esempio, quali sono i difetti di una certa auto o



Dalla Sentiment Analysis alla Content Analysis

Si possono anche avere informazioni simili a quelle delle indagini campionarie classiche

ad esempio, quali sono i difetti di una certa auto o

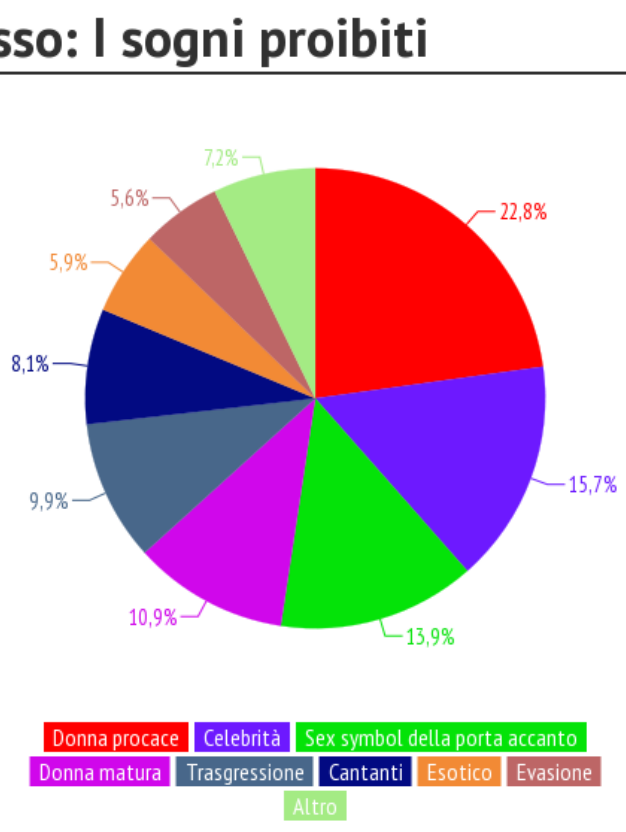


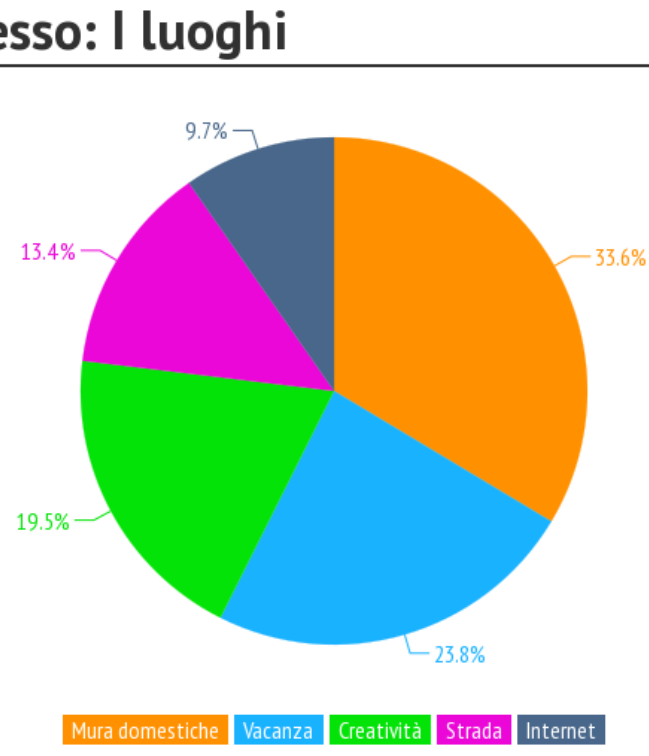
repubblica.it/economia/2014/07/15/news/voices_from_the_blogs_80_euro_renzi-91621990/

... e in rete ci si sente più liberi di esprimersi

voicesfromtheblogs.com/2014/07/23/sexandthetweet-summertime-i-sogni-erotici-degli-italiani-s

#SexandtheTweet: Summertime! **I sogni erotici degli italiani** (sotto l'ombrellone)

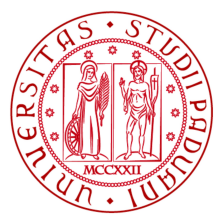




Dear Jack su FaceBook

(breve e semplice) analisi degli ultimi 5000 post della pagina `www.facebook.com/dearjackrock`

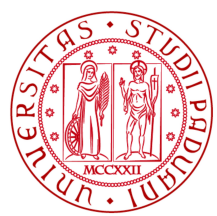
Dear Jack su FaceBook: più femmine o più maschi?



Dear Jack su FaceBook: più femmine o più maschi?

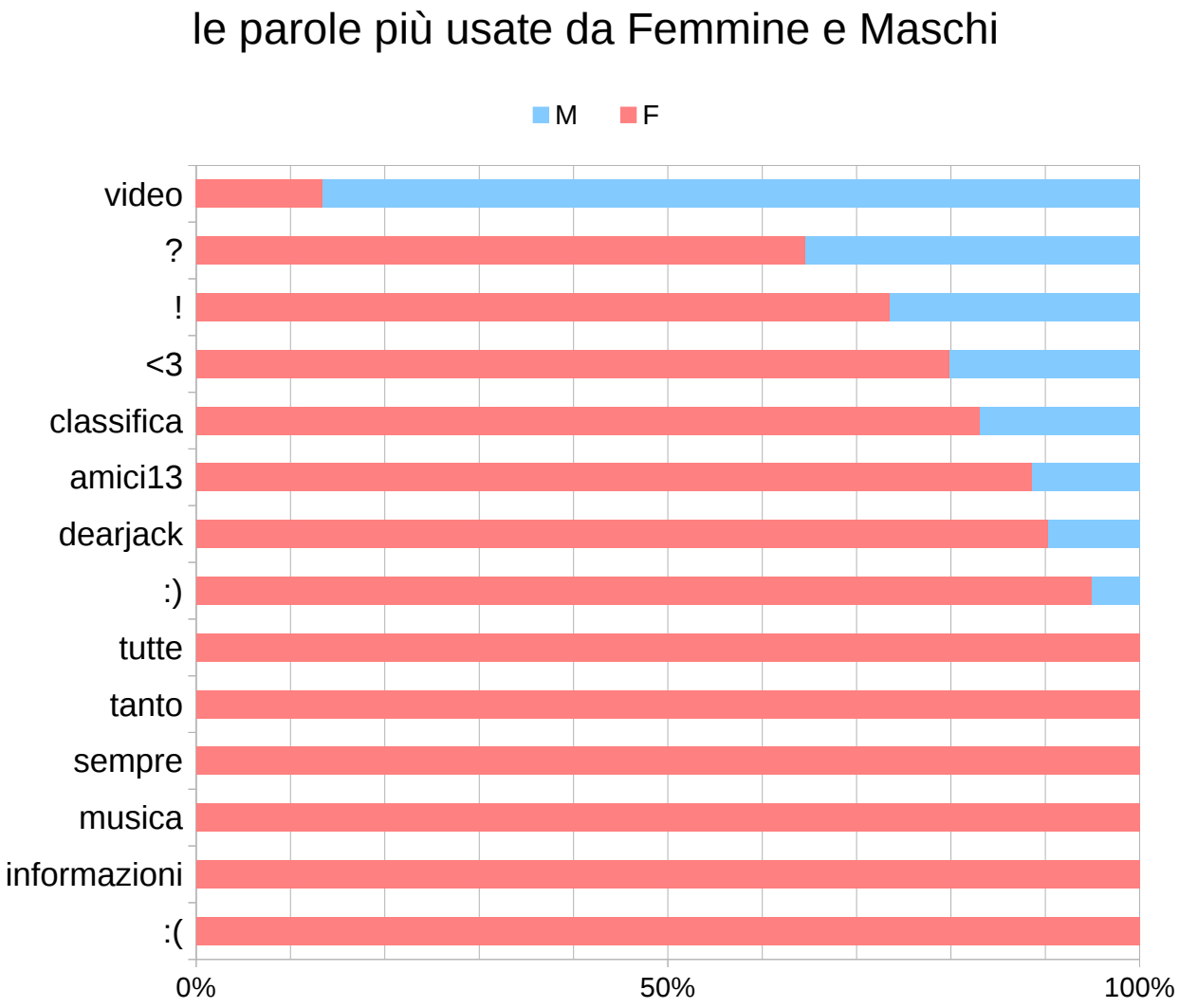
Il genere degli utenti è quasi sempre pubblico

	Femmine	Maschi	Totale
Conteggi	1266	635	1901
%	66.60	33.40	100.00

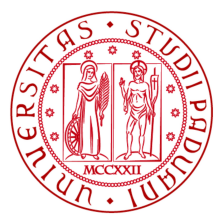


Le parole più 'caratteristiche' usate dai due generi

Posso facilmente raccogliere i post di ognuno di loro e cercare tra le parole usate molto da un genere e poco dall'altro:



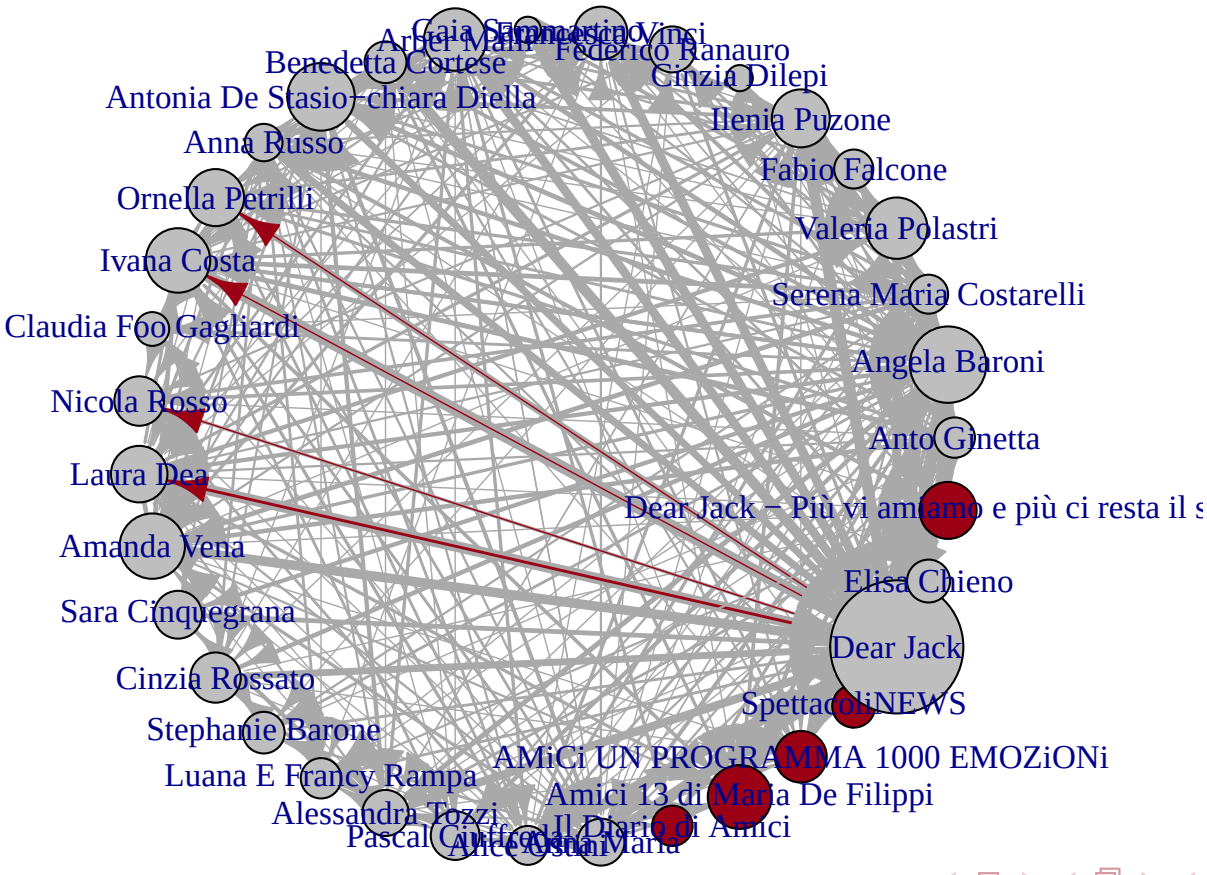
Se 'posto' chi mi risponde? Un'analisi di rete



Se 'posto' chi mi risponde? Un'analisi di rete

Spigoli: Post ← (commento o like)

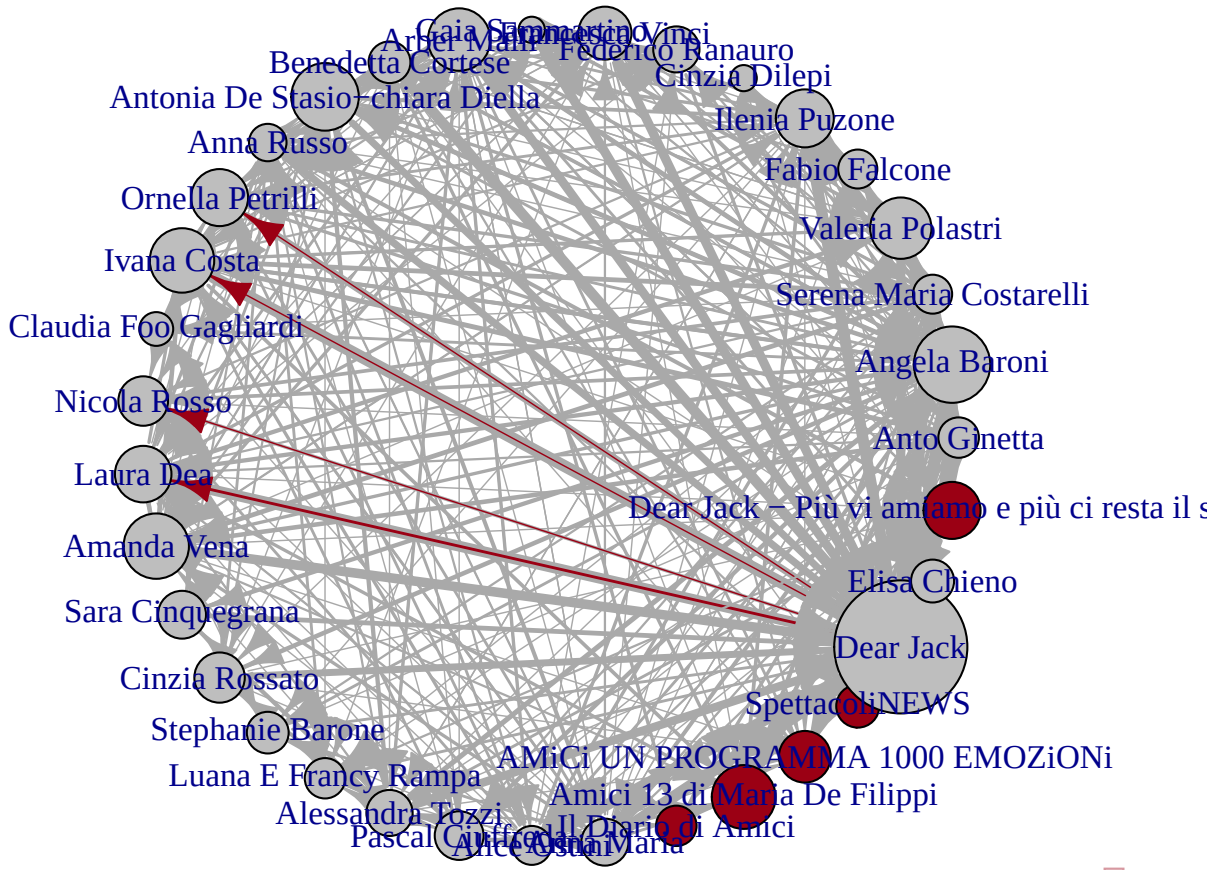
spessore cresce con il numero di contatti



Se 'posto' chi mi risponde? Un'analisi di rete

(Quasi) tutti hanno risposto a DearJack

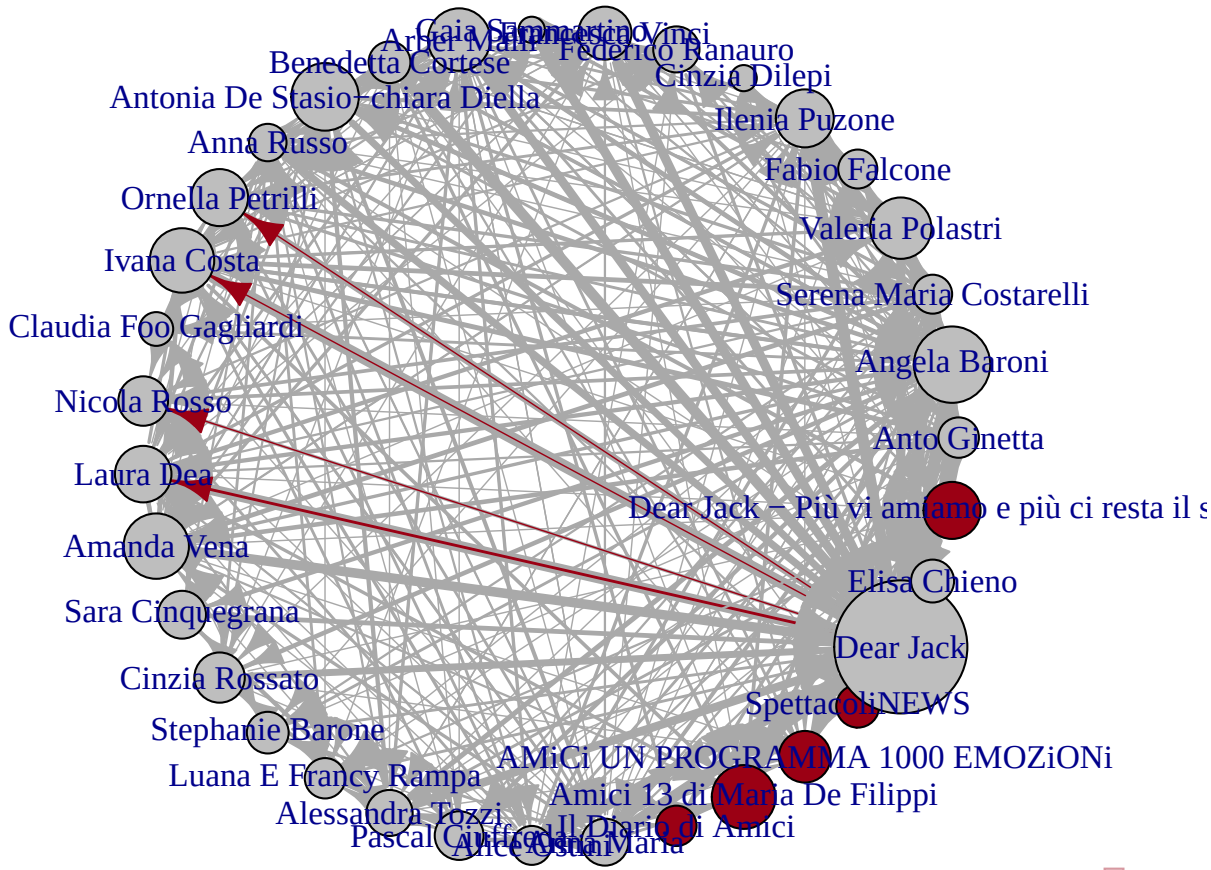
(= Spigoli spessi e puntati su DearJack)



Se 'posto' chi mi risponde? Un'analisi di rete

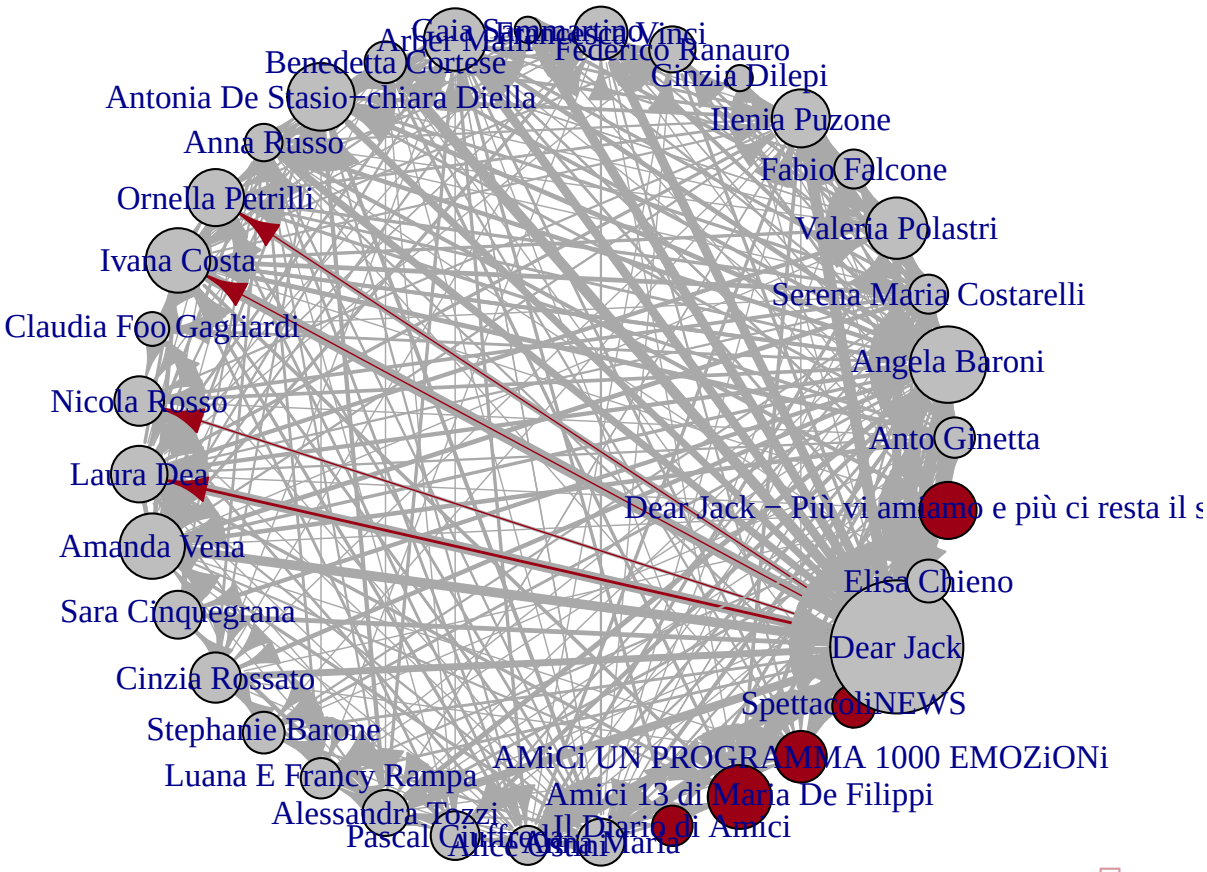
DearJack risponde a pochi

(= pochi e sottili Spigoli partono da DearJack)



Se 'posto' chi mi risponde? Un'analisi di rete

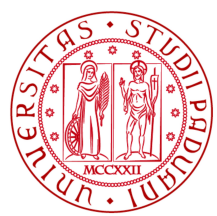
Gli utenti non persone fisiche (in rosso)
non scrivono a DearJack (sono lì per fare altro?)



... e veniamo a noi :)

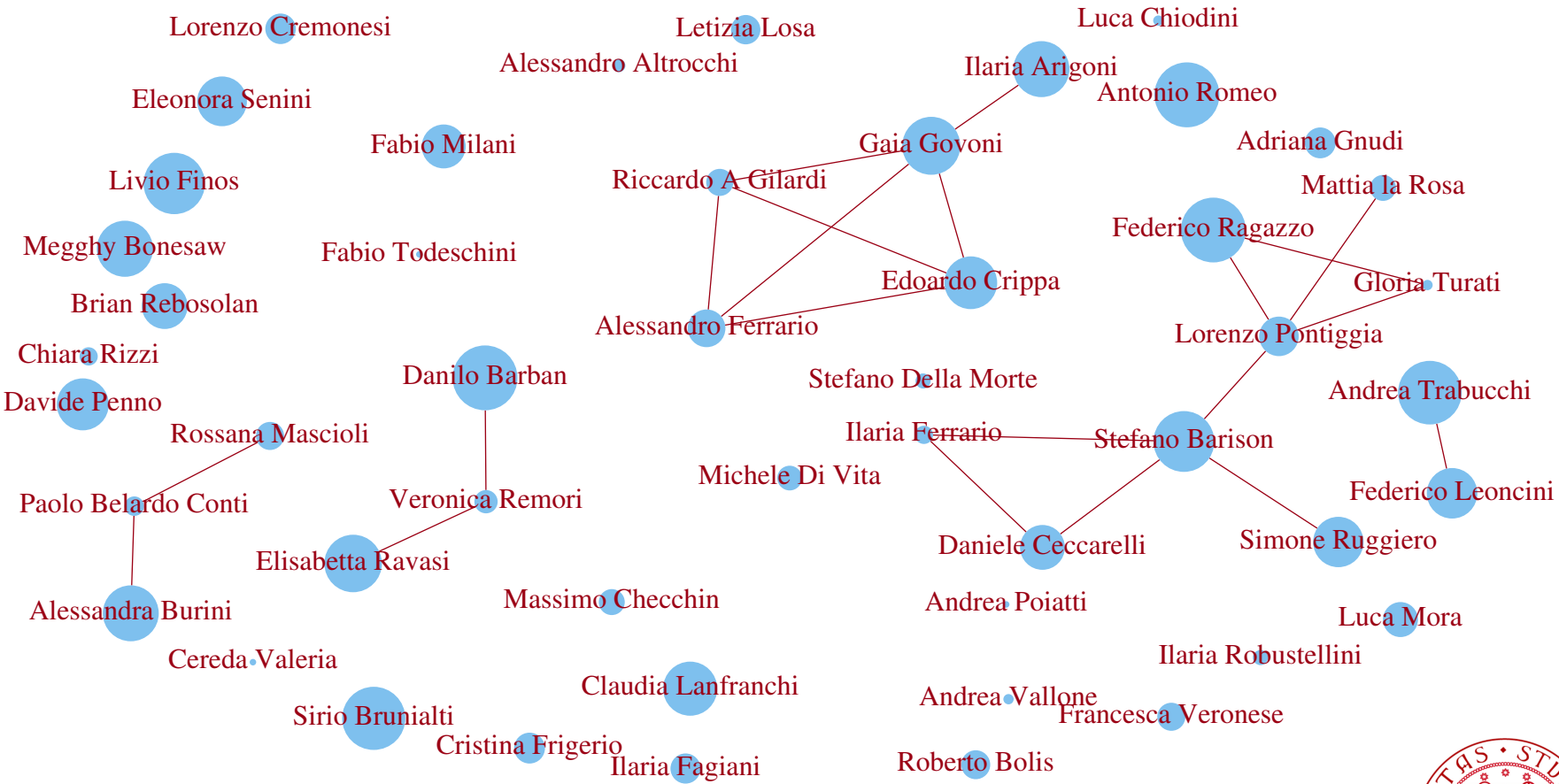
facebook.com/matescienze.summerschool

Cosa so di voi?



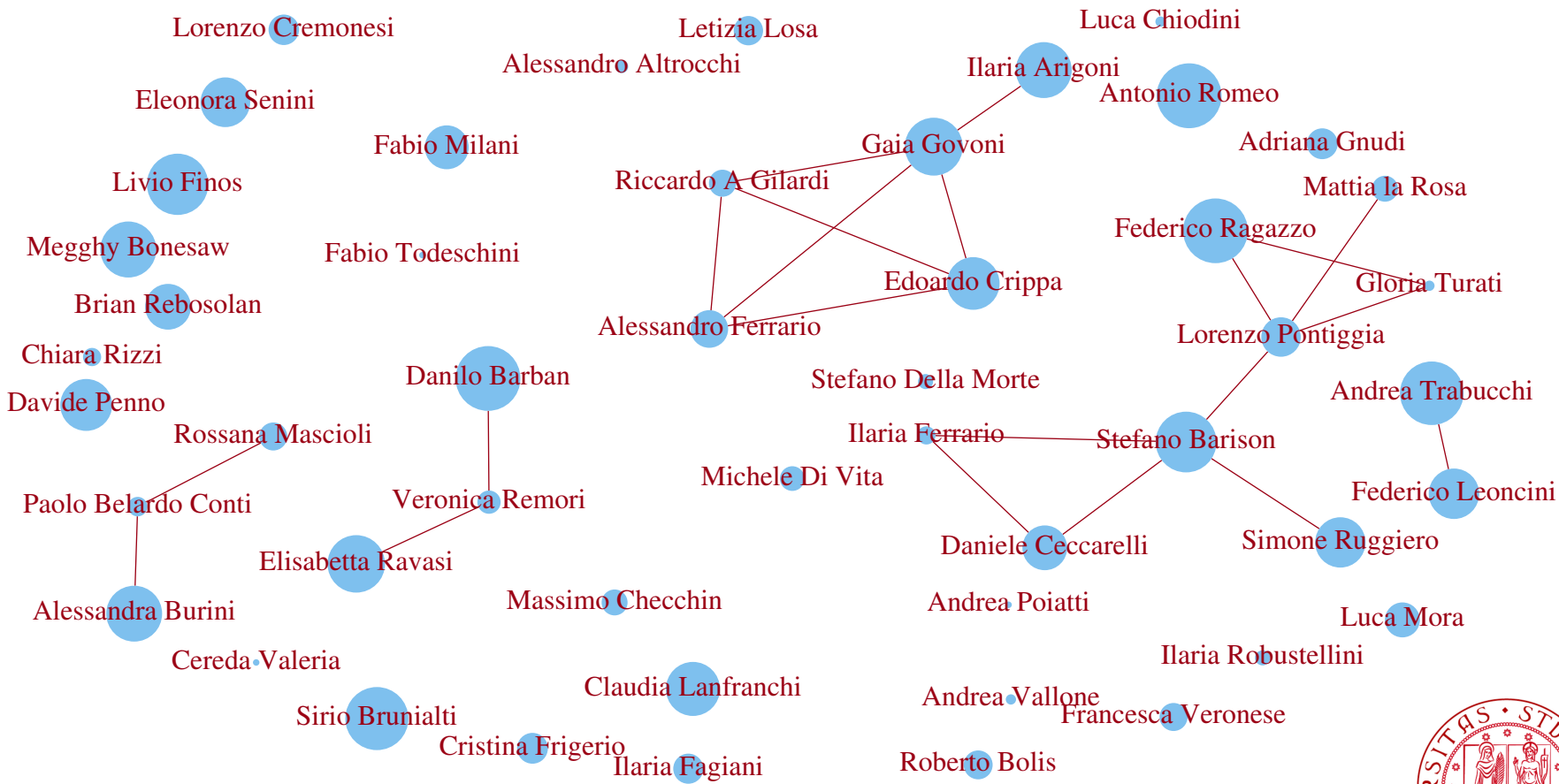
chi parla con chi

Vertici: diametro cresce con il numero di post (max 1000)



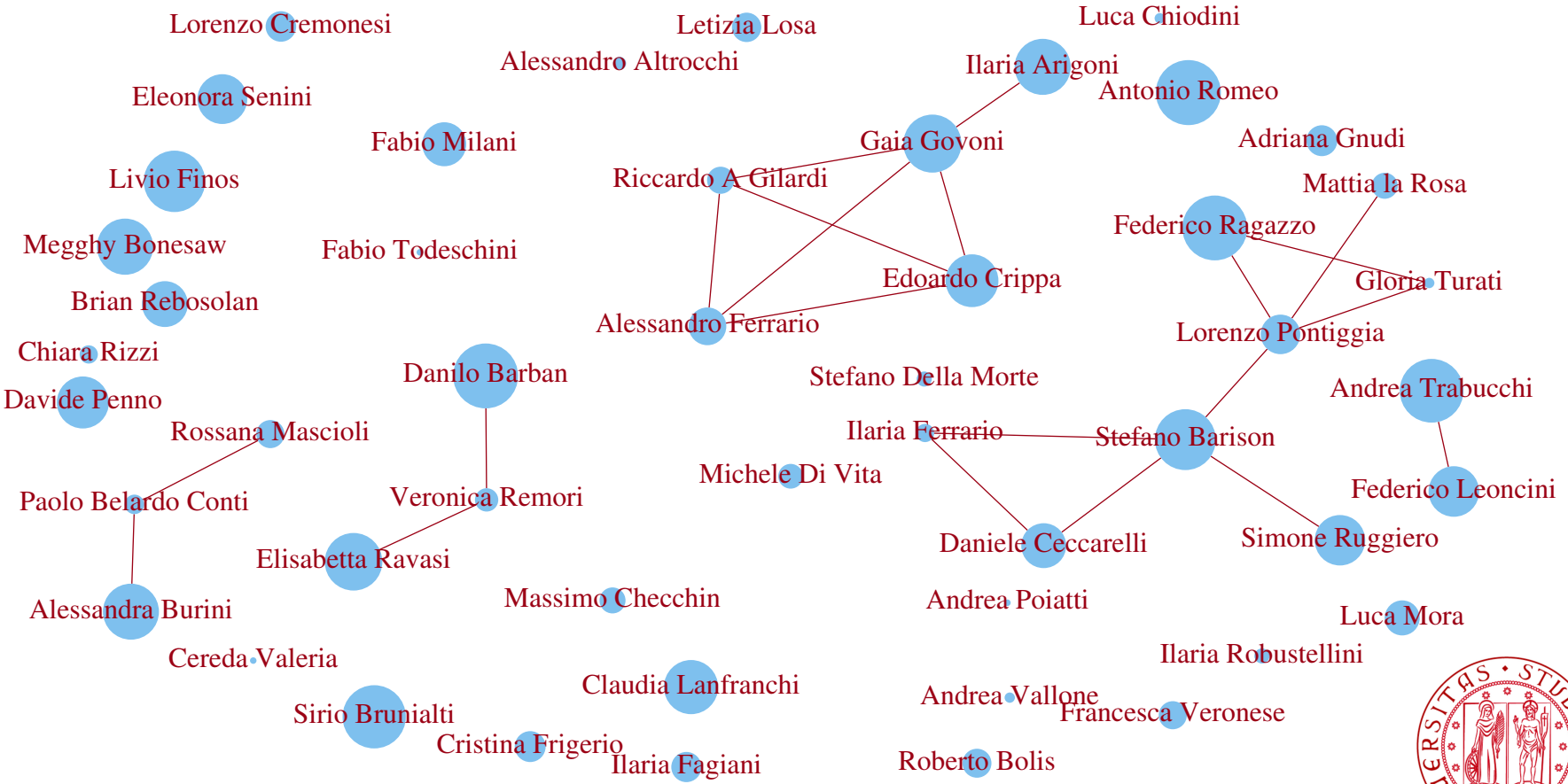
chi parla con chi

Spigoli: esiste un collegamento se un utente i) posta nel 'wall' dell'altro o ii) commento/like



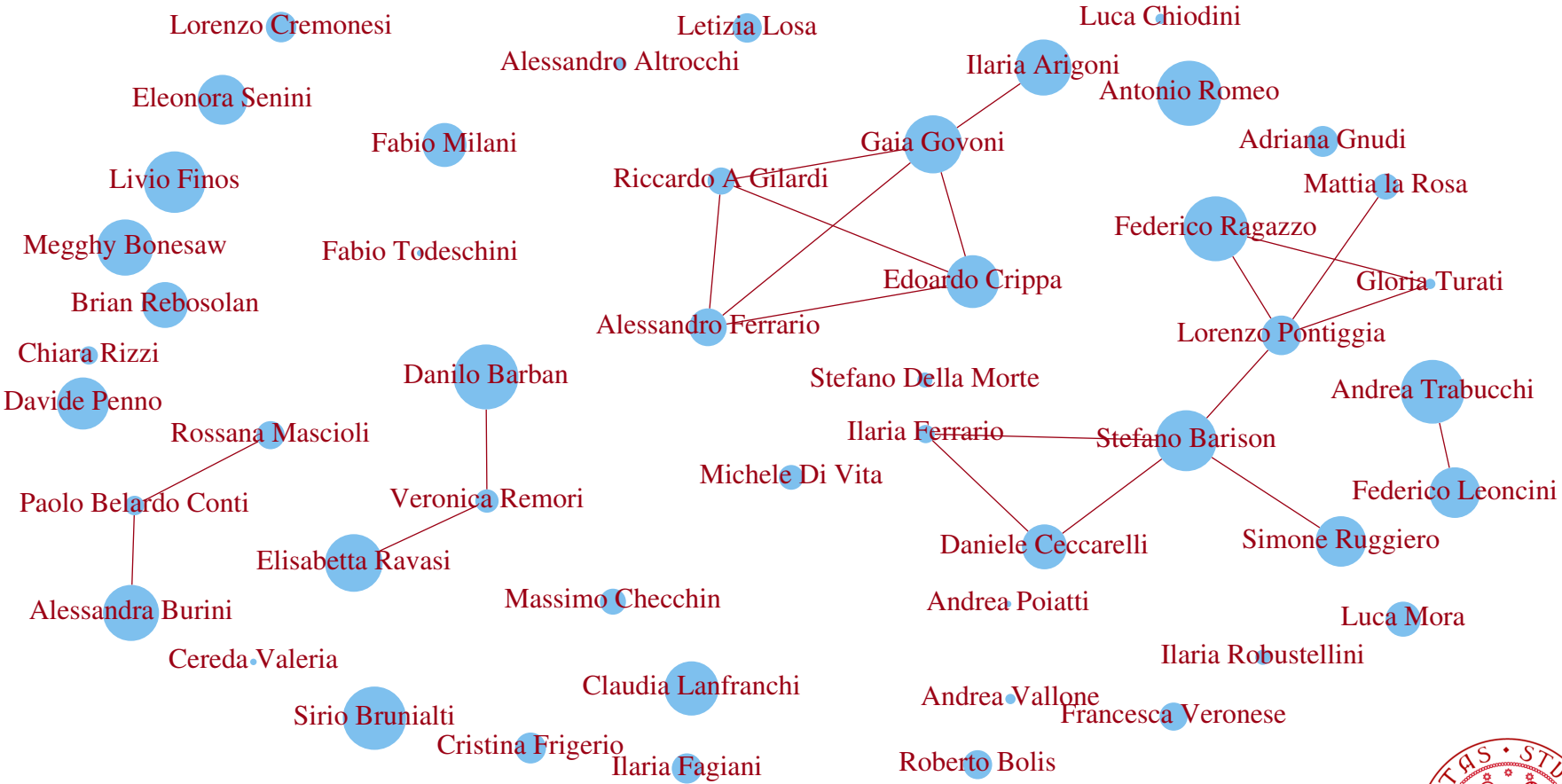
chi parla con chi

La rete è molto diversa da quella dei DearJack
(utenti molto più slegati perché non si conoscono e/o non hanno grande motivazione a condividere)



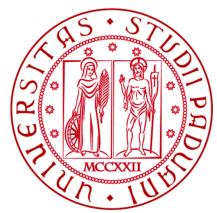
chi parla con chi

esistono però alcuni gruppi di amici (...)



Da dove veniamo?[2]

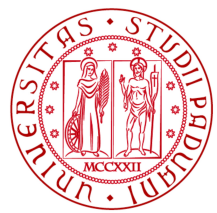
	Conteggi	%
Bergamo	16	32.65
Monza-Brianza	6	12.24
Lecco	5	10.20
Varese	5	10.20
Milano	4	8.16
NonDichiara	4	8.16
Lodi	3	6.12
Sondrio	3	6.12
Estero	1	2.04
Pordenone	1	2.04
Verona	1	2.04
Totale	49	100.00



[2] campi hometown o location

Impegni sentimentali..

	Conteggi	%
Single	13	59.09
In a relationship	3	13.64
Married	3	13.64
Engaged	2	9.09
It's complicated	1	4.54
Totale	22	100.00

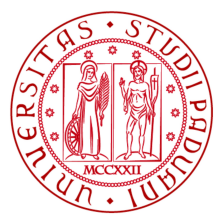


Studenti vs prof

prof = data di nascita dichiarata $\leq$ 01/01/1990

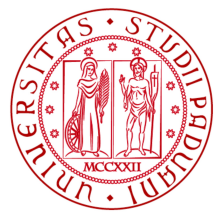
	studente	prof	?	Totale
Conteggi	37	5	7	49
%	75.51	10.20	14.29	100.00

Chi sono i 7 personaggi misteriosi? **studenti** o **prof**?



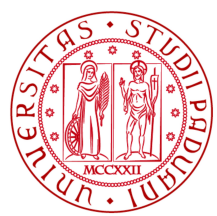
Ruoli e impegni

Conteggi	studente	prof	?
Impegnati[3]	5	3	0
NonDichiara	21	2	4
Single[4]	11	0	3
Totale	37	5	7



Ruoli e impegni

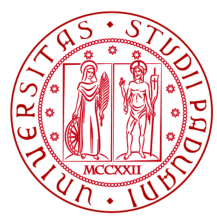
%	studente	prof	?
Impegnati	13.51	60.00	0.00
NonDichiara	56.76	40.00	57.14
Single	29.73	0.00	42.86
Totale	100.00	100.00	100.00



Ruoli e impegni

%	studente	prof	?
Impegnati	13.51	60.00	0.00
NonDichiara	56.76	40.00	57.14
Single	29.73	0.00	42.86
Totale	100.00	100.00	100.00

P(**Single**| **prof**)=0.00
P(**Single**| **studente**)=0.2973
se sapessi che un **?** è **Single** scommetterei che più probabilmente è uno **studente**.



Una predizione per i 7 personaggi misteriosi

- Raccolgo tutte le informazioni che ho su di voi (**profilo**) es: numero di amici, genere, numero di post, numero di '?' e di '!', numero di ':)', di likes, di parole usate nei post, etc.

[3] oltre alle probabilità condizionate entra in gioco anche il teorema di Bayes, ma non ne faremo accenno qui.

Una predizione per i 7 personaggi misteriosi

- Raccolgo tutte le informazioni che ho su di voi (**profilo**) es: numero di amici, genere, numero di post, numero di '?' e di '!', numero di ':)', di likes, di parole usate nei post, etc.
- Stimo P(**profilo X** | **studente**) e P(**profilo X** | **prof**) sulla base dei dati noti (gli altri utenti di cui so la data di nascita) Stima tramite **analisi discriminante** o altri metodi più complessi.

[3] oltre alle probabilità condizionate entra in gioco anche il teorema di Bayes ma non ne faremo accenno qui.

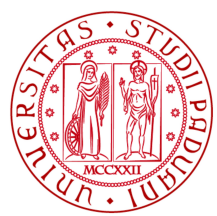
Una predizione per i 7 personaggi misteriosi

- Raccolgo tutte le informazioni che ho su di voi (**profilo**) es: numero di amici, genere, numero di post, numero di '?' e di '!', numero di ':)', di likes, di parole usate nei post, etc.
- Stimo P(**profilo X** | **studente**) e P(**profilo X** | **prof**) sulla base dei dati noti (gli altri utenti di cui so la data di nascita) Stima tramite **analisi discriminante** o altri metodi più complessi.
- se P(**profilo X** | **studente**) $>$ P(**profilo X** | **prof**)[3] scommetto che il **profilo X** appartiene ad uno **studente**.

[3] oltre alle probabilità condizionate entra in gioco anche il teorema di Bayes ma non ne faremo accenno qui.

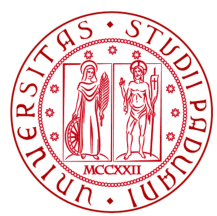
Le mie scommesse sui 7 personaggi misteriosi

Utente FB	Scommetto che è:
Claudia Lanfranchi	studente
Riccardo A Gilardi	studente
Luca Chiodini	studente
Alessandro Ferrario	studente
Alessandra Burini	studente
Stefano Della Morte	studente
Lorenzo Cremonesi	studente



Statistica + Informatica + Fantasia = Data Science

La faccia della statistica sta cambiando
Il volume e la qualità dei dati sono cambiati, anche i metodi matematico/statistici stanno cambiando.



Statistica + Informatica + Fantasia = Data Science

La faccia della statistica sta cambiando
Il volume e la qualità dei dati sono cambiati, anche i metodi matematico/statistici stanno cambiando.

Ci sono miniere di dati imbrigliate nella rete,
servono strumenti matematico/statistici (nuovi e vecchi) per distillare le informazioni
... e la fantasia per farne una storia da raccontare.