Measuring Psychospread: idiosyncratic happiness of Italian citizens through the analysis of their tweets

---

Canova L. (Enrico Mattei School, enicorporateuniversity)
Curini L. (University of Milan)
Iacus S. (University of Milan)

FEEM, April 2nd

Presentation Agenda

- Introduction
- Literature
- Data and Methodology
- Results
- Conclusion

Introduction

• **Research question**: can we measure idiosyncratic happiness through the analysis of tweets posted on the web?

• Can be useful to disentangle idiosyncratic happiness from a more structural concept for policy purpose?

Literature

Literature is large on the measurement and definition of happiness
Empirical studies:

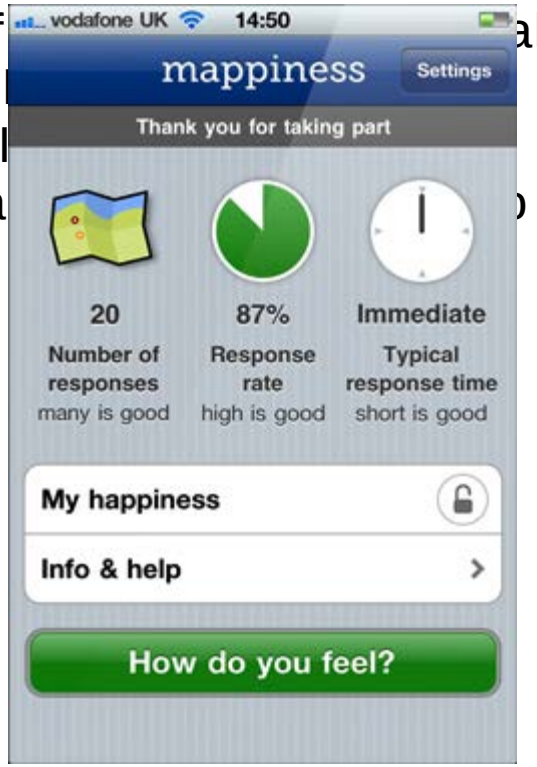
-**subjective well-being indicators** (self-reported happiness) [Veenhoven (1994); Lyubomirsky and Lepper (1999)]
-**DRM**: Daily Reconstruction Method [Kahneman (2004)]

Subjective well-being indicators
Robust correlation with health outcomes [Cohen et al., 2003] but…

Many problems arising from their use because of context dependence [Schwarz, 1987; Deaton (2013)]

Experience Sampling Me

•Inf als' well-b l time in t
•Kill
•Ma (2010)



Sentiment Analysis through Social Networks

Twitternomics:

-Bollen et al. (2011): Twitter used to predict stock prices

-Dodds et al. (2011) use Twitter to estimate happiness in United States. They use Amazon's Mechanical Turk:

- It is a website where volunteers are paid to rate the level of happiness of ten thousand words of English. The average score is then used to codify more than 60 millions of tweets from all the world

Data and Methodology: our ap

- Download posts from Twitter, which provides a classification of **happy/unhappy tweets** based on text emoticons used
- 2 steps procedure following Hopkins and King (2010):

-Use this sub-sample of tweets as a training set (first step)
-Extend classification of tweets (using an automated statistical analysis) to the entire population of Italian tweets at the provincial level

Construct an iHappy Index:

$$iHappy = \frac{(n.\ of\ happy\ tweets)}{(n.\ of\ happy \wedge unhappy\ tweets)} \times 100\%$$

Data

- **43 millions of tweets in 2012**
- **200 thousands** tweets posted daily
- Unit of analysis: **110 provinces (using ge**

•We derive an **happiness level** aggregated at provincial level

Pros and cons

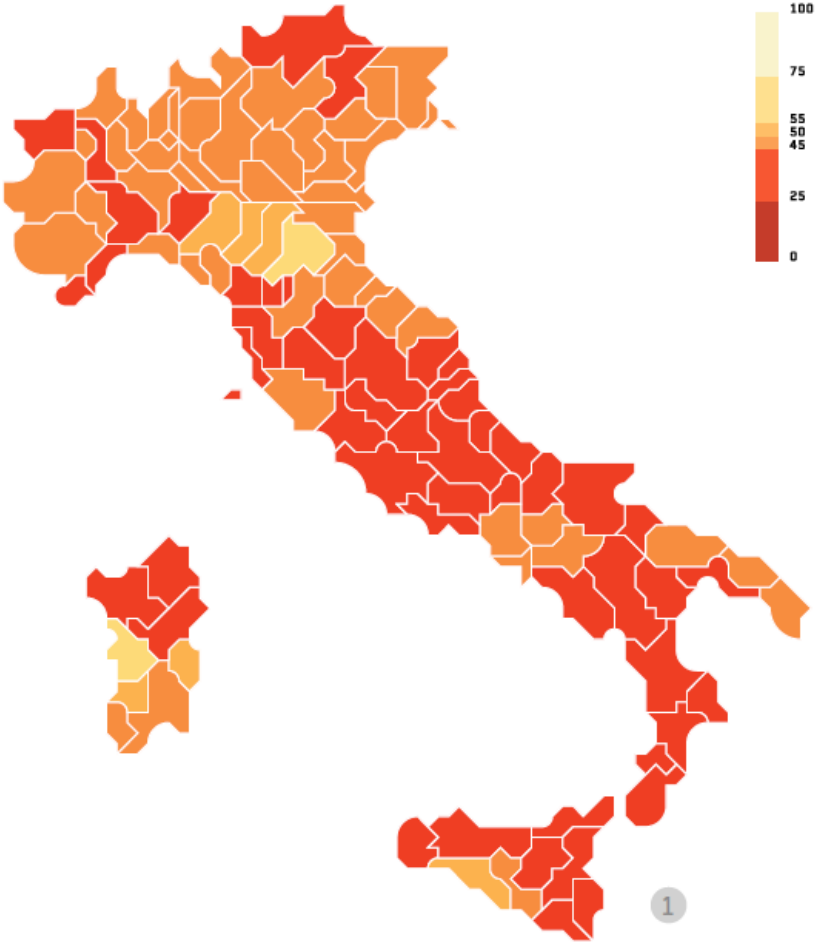
•We don’t use any questionnaire
•Exponential growth of Twitter use:

-third highest ranking social media, behind Facebook and MySpace, with a relevant growth from 2009 (when it ranked 22nd)
- in Italy, in 2013, over 4 millions of users active on Twitter, more than doubling in less than one year

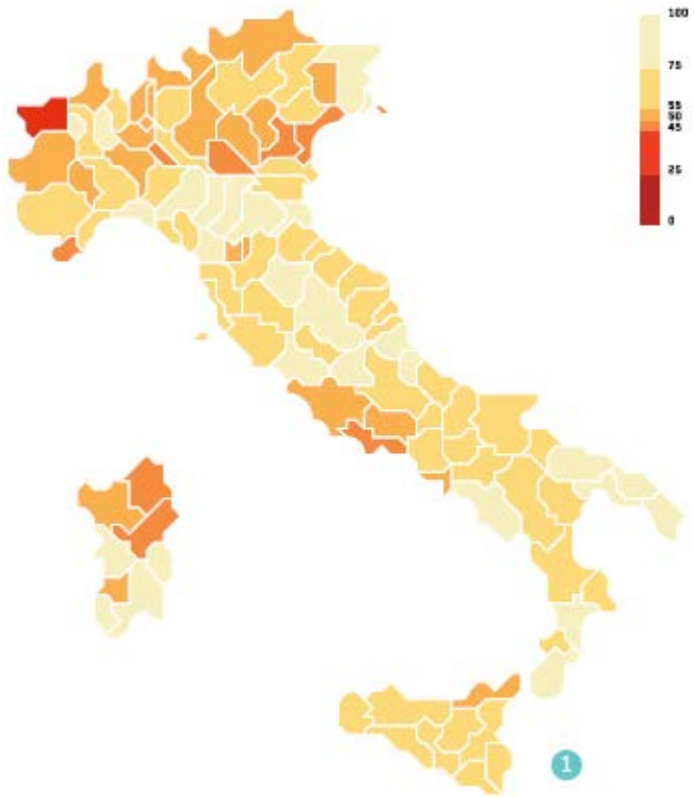
•Issue of **representativeness**:

- the sample of tweets analyzed cannot be representative of the whole population of Italian citizens
- Socio-economic characteristics of people having access to the web are different from those of the entire cohort of citizens (Tjong Kim Sang and Bos, 2012).

Some descriptive evidence

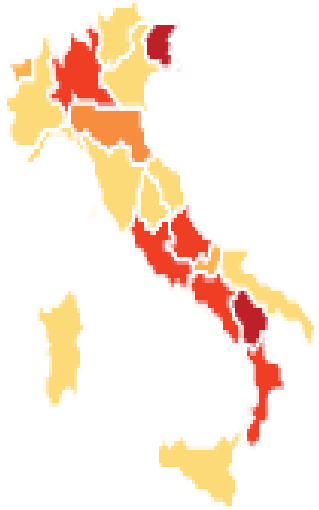


- **Avg value of iHappy Index**: 45,6%
- **Nr of days with iHappy above average**: 151
- **Nr of days with iHappy above 50%:** 111



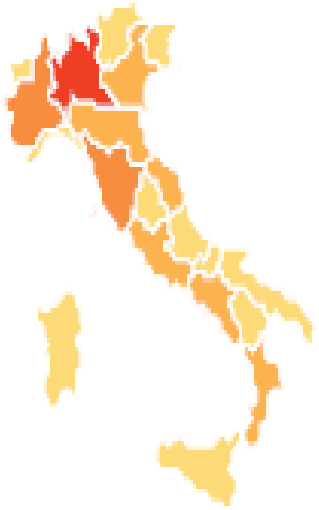
- **Avg value of iHappy Index**: 60.3%
- **Nr of days with iHappy above average**: 221
- **Nr of days with iHappy above 50%:** 310

Happiness peak: Balotelli



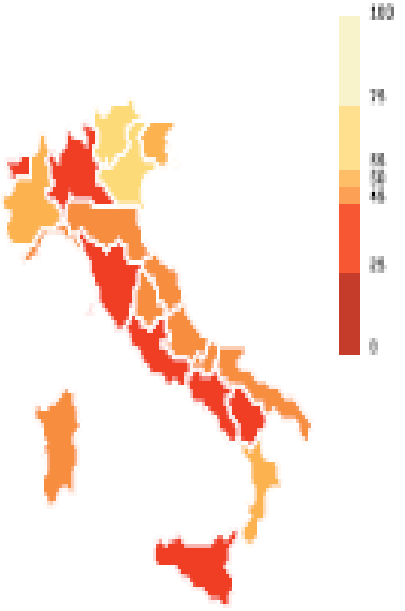
28 GIUGNO

Before goal
44%



29 GIUGNO

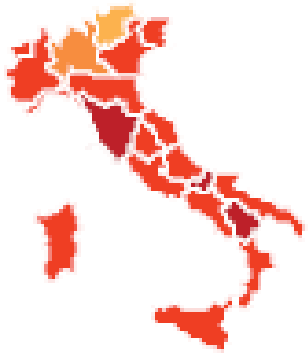
After the goal
51%



30 GIUGNO

Happiness gets back
45%

Picchi di felicità (1): Vittoria di Obama

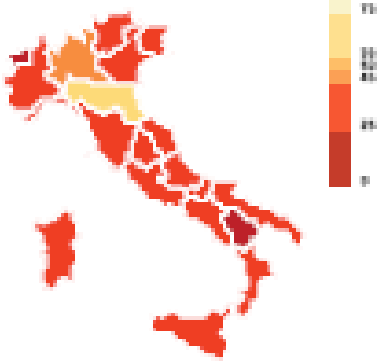


5 NOVEMBRE

Indice di felicità nell'ultimo giorno di voto
$iHappy_T$ = 36%

6 NOVEMBRE

L'euforia durante la notte dei risultati elettorali
$iHappy_T$ = 63%



7 NOVEMBRE

L'euforia è già rientrata il giorno successivo
$iHappy_T$ = 38%

Picchi di felicità (3): La notte prima della fine del mondo

19 DICEMBRE

Due giorni prima
iHappy = 42,0%

20 DICEMBRE

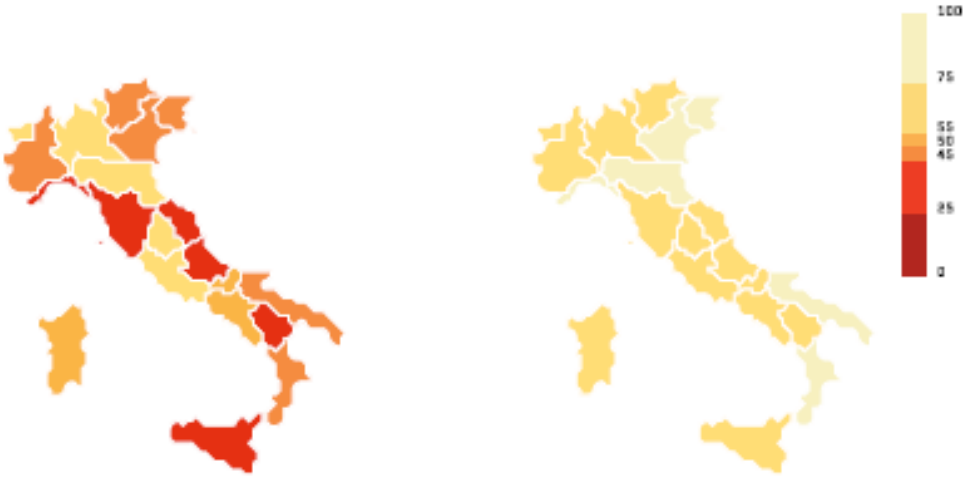
La notte prima della fine del mondo
iHappy = 64,5%

21 DICEMBRE

Il mondo alla fine va avanti
iHappy = 37,4%

2013

Il giorno dopo le elezioni politiche iHappy: +20,6%



25 FEBBRAIO 2013

iHappy 50,1%

26 FEBBRAIO 2013

iHappy 70,7%

La nave Concordia sull'isola del Giglio viene raddrizzata grazie ad una imponente operazione ingegnieristica durata 19 ore
iHappy: +18,5%

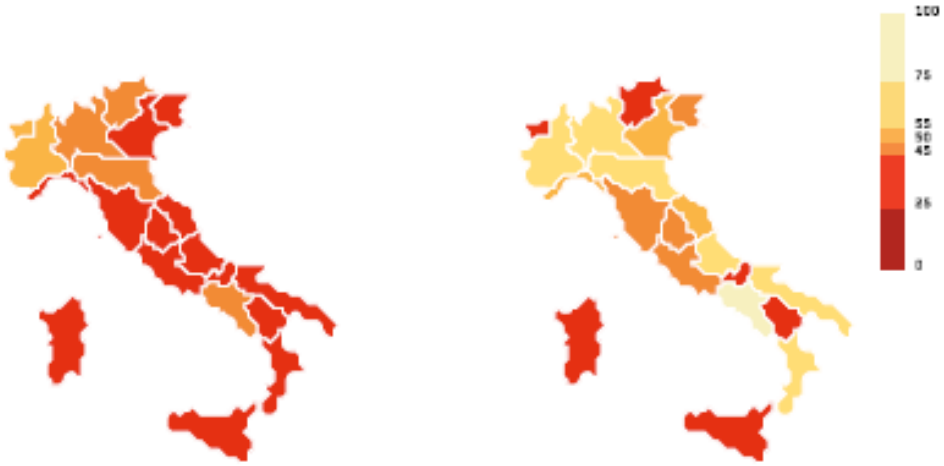
16 SETTEMBRE 2013

iHappy 52,6%

17 SETTEMBRE 2013

iHappy 71,1%

Felicità per l'anno nuovo e la fine del 2012
iHappy: +18%



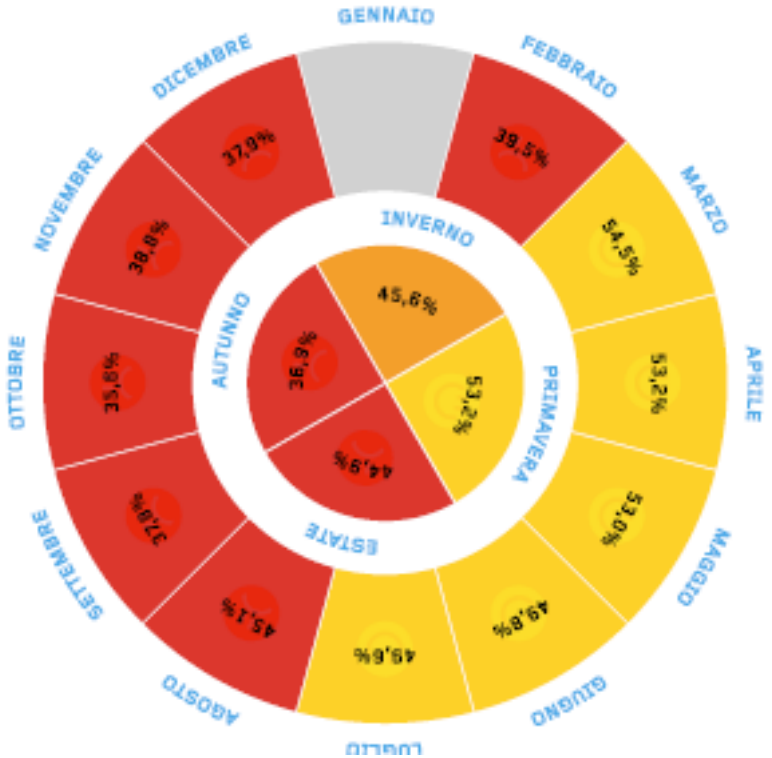
31 DICEMBRE 2013

iHappy 40,2%

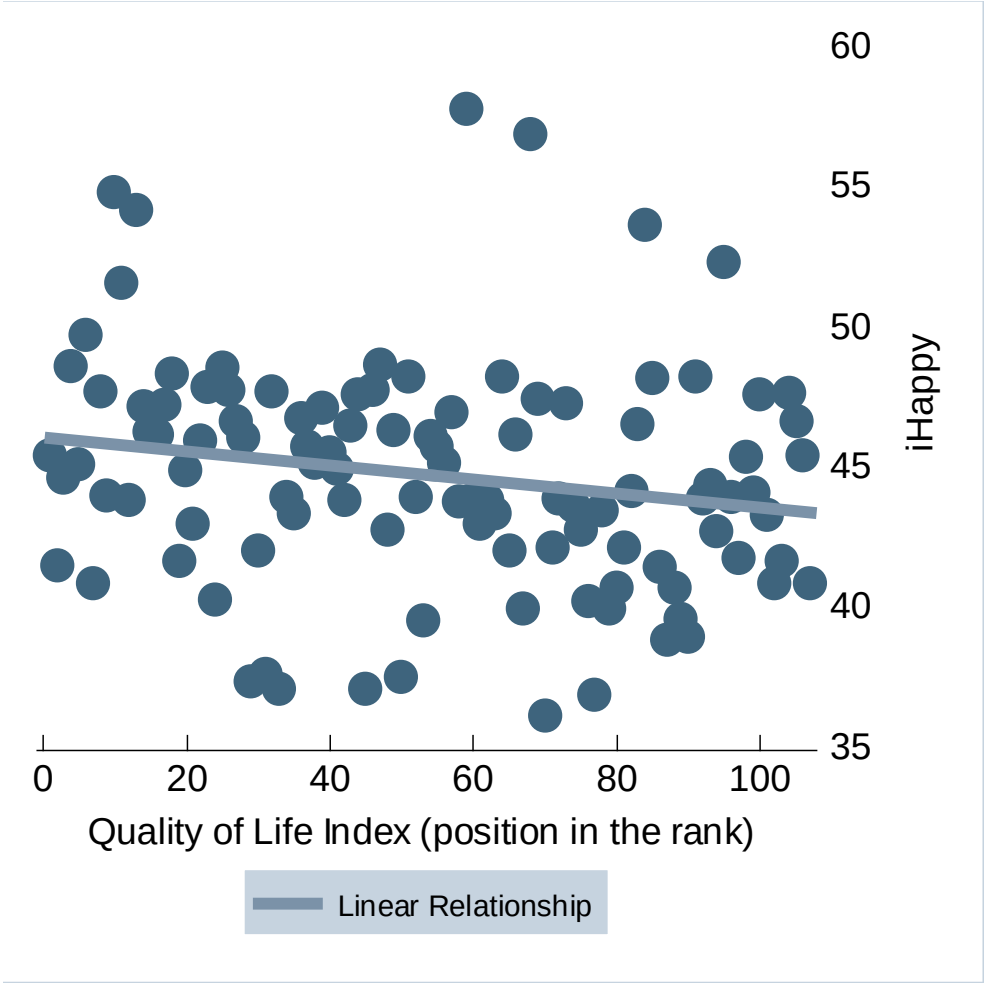
1 GENNAIO 2013

iHappy 58,2%

●	1	Oristano	56,6%
●	2	Bologna	56,1%
●	3	Modena	55,0%
●	4	Ogliastra	54,1%
●	5	Medio Campidano	54,1%
●	6	Agrigento	53,6%
●	7	Reggio Emilia	52,1%
●	8	Parma	50,6%
●	9	Pordenone	49,9%
●	10	Forlì	49,8%



Correlation between iHappy and Quality of Life Sole 24 Ore



-0.19 but robust and significant (1%)

Econometric analysis

Dynamic OLS using iHappy as dependent variable
Set of correlates:

•**emotional variables**: the set of correlates includes lag of iHappy index in the 3 previous days; dummy variables for a number of feasts: Valentine's day; dad and mother's feast day; Christmas and New Year's Eve; August, 15th
•time variables: we consider the day of the week, together with the month of the year and season

•**meteorological regressors**: controls take into consideration temperature of the day, weather (sun, snow or rain) and interaction terms

•**socio-demographic characteristics and institutional variables**: we control for the percentage of singles at the provincial level; gender distribution; population and squared population; latitude; altitude; birth rate; quality of infrastructures available for leisure; distribution of enterprises; public order

•**economic explanatory variables**: the set of our regressors includes both static elements, like log of average provincial income, and idiosyncratic ones, which are dummy variables controlling for payday; tax return day; lag of spread level

	est1	
Lag of iHappy (value of the previous day)	0.207***	(0.008)
Lag of iHappy (value of 2 days before)	0.122***	(0.009)
Lag of iHappy (value of 3 days before)	0.135***	(0.008)
Payday	-3.299***	(0.347)
Lag of payday	1.301***	(0.360)
Tax return day	-2.814***	(0.490)
Temperature	-0.489***	(0.072)
Winter	-1.825*	(0.821)
Winter*temperature	0.341***	(0.059)
Summer	-19.698***	(5.702)
Summer*temperature	1.583**	(0.482)
Squared temperature	0.010***	(0.002)
Summer*squared temperature	-0.029**	(0.010)
Spring	-0.961	(1.107)
Spring*temperature	0.252***	(0.063)
Rain	-0.425*	(0.185)
Snow	-0.323	(0.535)
Bank holidays	0.040	(0.328)
Womens' day	2.657***	(0.615)
Mother's day	8.658***	(0.700)
Valentine's day	2.899**	(1.012)
New Year's Eve	0.859	(1.076)
August, 15th	5.517***	(1.619)
Christmas	14.987***	(1.577)
Father's Day	2.588***	(0.695)
Thursday	1.316***	(0.267)
Monday	1.004**	(0.312)
Tuesday	2.415***	(0.319)
Wednesday	-0.908**	(0.300)
Saturday	2.009***	(0.271)
Friday	0.318	(0.286)
Lag of spread value	-0.028***	(0.002)
April	3.088***	(0.597)
December	-7.732***	(0.994)
February	-5.311***	(0.809)
June	2.724***	(0.369)
July	5.295***	(0.368)
May	3.629***	(0.507)
March	1.581*	(0.696)
November	-4.222***	(0.777)
October	-5.184***	(0.642)
September	-3.603***	(0.608)
Latitude	0.082	(0.124)
Altitude	-0.001	(0.001)
Log of average income (provincial level)l	-3.213	(4.569)
Average age (provincial level)	1.565***	(0.416)
Percentage of singles	75.905***	(20.437)
Birth rate	1.021*	(0.479)
Population	0.000***	(0.000)
Squared population	-0.000***	(0.000)
Provincial capital	0.408	(0.414)
Public Order	0.004	(0.007)
Leisure infrastructures	0.036**	(0.011)
% companies over population	-14.642	(18.460)

Some results in detail

• **Lag of iHappy index** has a positive sign and it is statistical significant at 1% level: memory of happiness in the previous days generates a sort of persistent effect on the happiness of days after

• **average income** does not produce any statistically significant consequence, confirming economic literature on irrelevance of income level (C

• **Lag of payday** increases iH n the day of tax return, iHappy ind

• dependent variable is affect oints increase of of spread reduces he day after.

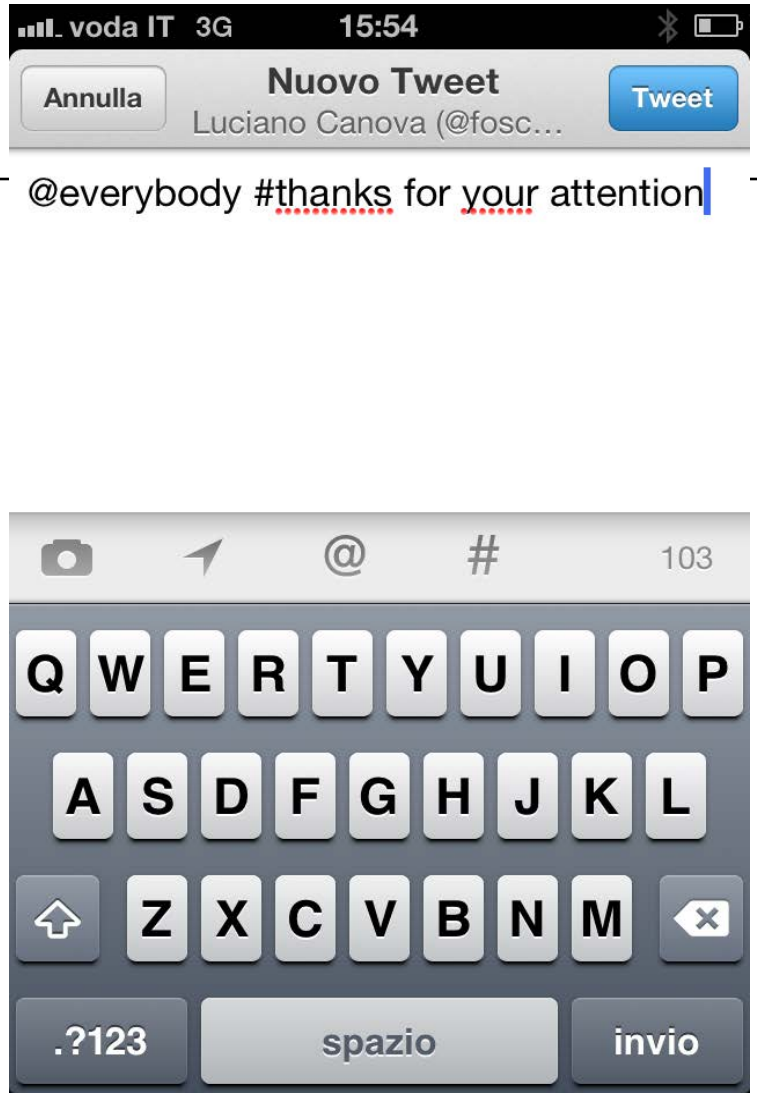
Conclusions

•We construct an aggregate level of happiness valid at provincial level (110 provinces) using Twitter
•Statistically robust evidence of idiosyncratic shocks produced by a set of events

• Happiness level seems to persist in the short run. There is a memory of events which lasts for some time
• Happiness level gets back to its original value in the long run (adaptation? Hedonic Treadmill)

• Possibility of using happiness indicator as proxies for quality of life: possible correlation with social capital indicators?
• Necessity to investigate a representative population but…

Very interesting approach to develop!


voda IT 3G
15:54
Annulla
Nuovo Tweet
Luciano Canova (@fosc…
Tweet
@everybody #thanks for your attention
@
#
103
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M
.?123
spazio
invio