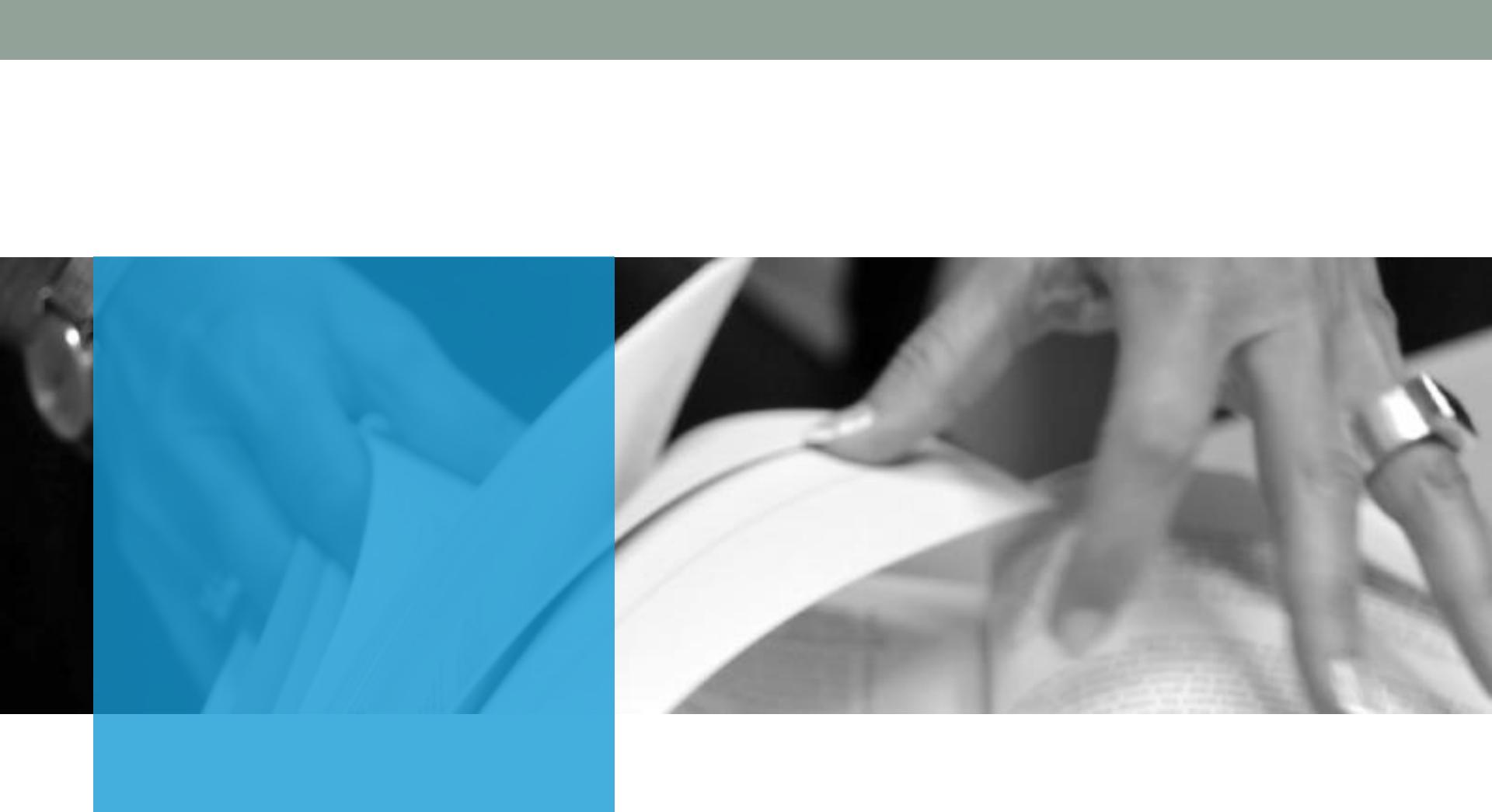


SENTIMENT ANALYSIS APPLICATIONS



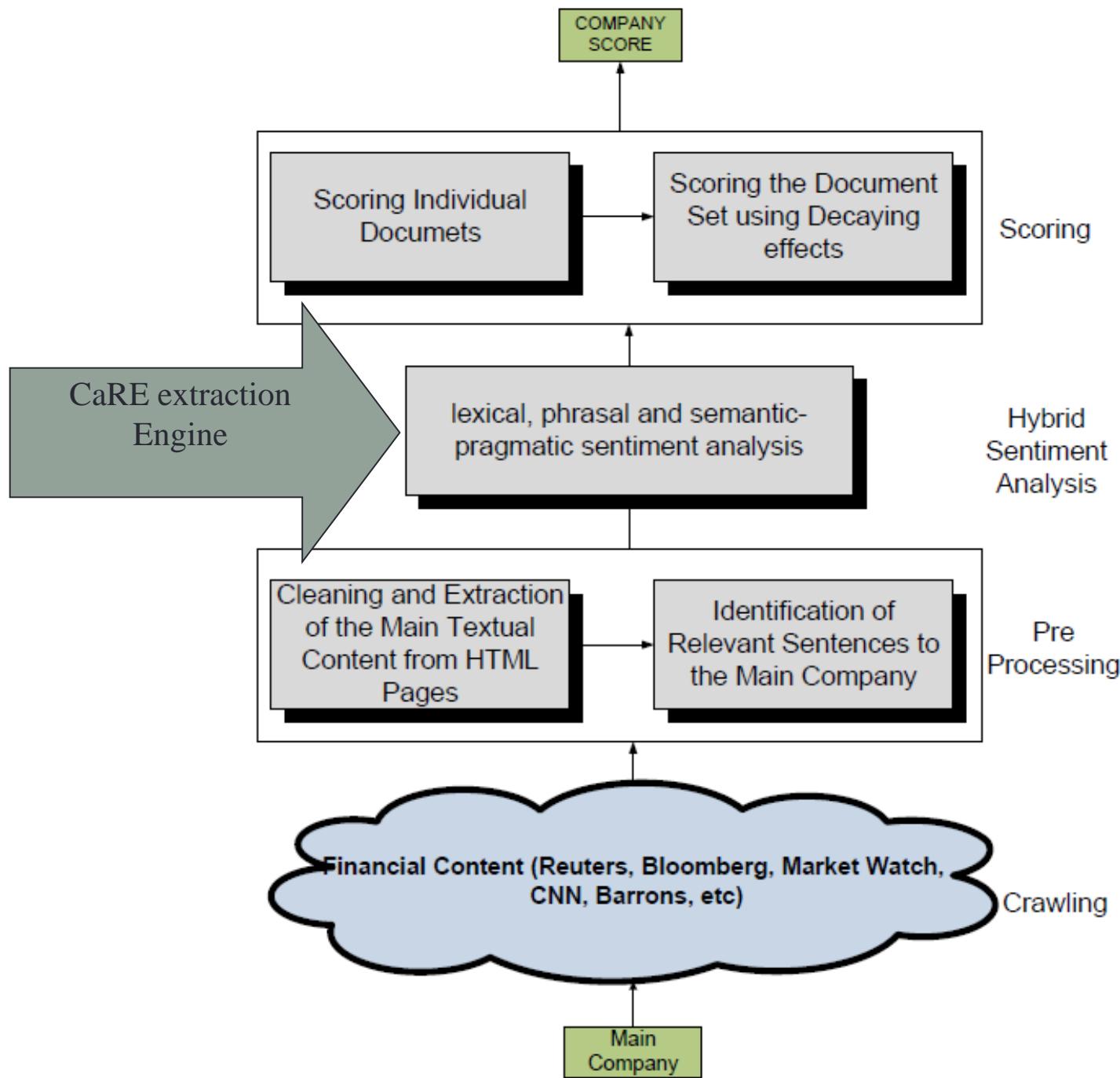
Sentiment Analysis of Stocks from News Sites

The Need for Event Based SA

*Toyota announces voluntary recall of their **highly successful top selling***

2010 model-year cars

- Phrase-level SA:
 - highly successful top selling ⇒ **positive**
 - Or at best neutral
 - Taking into account voluntary recall ⇒ **negative**
- Need to recognize the whole sentence as a “product recall” event!



Template Based Approach to Content Filtering

```
BusinessWeek
<BODY class="topNews">
  <DIV id="wrapper">
    <DIV id="header">
    <DIV id="columnLayout">
      <DIV id="column1">
        <DIV id="strapBox">
          <DIV id="storyBody">
            <H1>
            <H2>
            <P class="byline">
            <DIV id="inset">
              <P>
              <P>
              <P>
              <H3>
              <P>
              <P>
              <H3>
              <P>
              <P>
              <H3>
              <P>
              <P>
              <P>
              <P>
              <P>
              <P>
              <P>
              <P>
              <P class="tagline">
            </DIV>
            <DIV class="module">
            <DIV class="module">
              <DIV id="column2">
                <DIV id="footer">
```

The screenshot shows a news article from BusinessWeek titled "Why Subprime Lenders Are In Trouble". The article discusses how subprime mortgage standards were lowered in 2006. A sidebar on the right contains "STORY TOOLS" with links for commenting, emailing, printing, ordering a reprint, digg, and saving to del.icio.us. The page also features a navigation bar at the top and sections for most popular stories and multimedia.

BusinessWeek

keyword or company Search

Register | Sign In

OUR NEW LIFESTYLE CHANNEL

Current Issue Risk-Free Issues Subscribe Now

DIRTY DEEDS

HOME INVESTING COMPANIES TECHNOLOGY INNOVATION MANAGING SMALL BIZ B-SCHOOLS ASIA EUROPE LIFESTYLE MAGAZINE

Top News News Archive Special Reports Columnists Newsmaker Videos Newsletters The Debate Room Blogs

NEWS ANALYSIS March 2, 2007, 12:00AM EST

Why Subprime Lenders Are In Trouble

New analysis suggests that subprime lenders lowered their lending standards last year as they competed for business

by Peter Coy

As the subprime mortgage market goes into steep decline, threatening to drag the whole economy along with it, many people are wondering what could have gone so wrong so quickly. Until recently, after all, delinquency and foreclosure rates on subprime loans were reassuringly low.

The answer may lie in how the quality of these mortgages has changed over the years. While subprime is the term used for loans issued to people with poor credit, not all subprime loans are created equal. And the subprime loans that were originated in 2006 that are turning out to be shockingly weak.

Why 2006? What happened last year that caused credit quality to go into steep decline? Michael Youngblood, head of asset-backed securities research at Friedman, Billings, Ramsey Group ([FBR](#)) in Arlington, Va., has been poring over the data. He points out that there was a sudden but little-noticed shift in lenders' strategy that occurred at the end of 2005: Lenders went from competing for customers on price (by lowering rates) to competing for customers on easy terms (by lowering lending standards).

UNDISCLOSED CHANGE

The change was little-noticed, says Youngblood, because the lenders actively denied it. "To my disappointment as a long-time analyst," says Youngblood, the major lenders insisted that they had not lowered their credit standards long after they had begun to do so. "I met with all of the public subprime lenders in early June...to a man they swore they had maintained 2005

MOST POPULAR STORIES

Read E-mailed Discussed

1. Steady Income in Unsteady Times
2. Costco Starts a Barroom Brawl
3. Ten Things That Won't Happen in 2008
4. World Wide Web: Land of Free Stuff
5. New Home Sales: A Stunning Slide

RSS Feed: Most Read Stories

MOST POPULAR MULTIMEDIA

Slide Shows

1. The Most Important Products of 2007 - BusinessWeek
2. 101 Best Web Freebies - BusinessWeek
3. Asia's Oldest Bosses - BusinessWeek
4. The Best and Worst Housing Markets 2007 - BusinessWeek
5. Toyota's Hybrid Pickup Concept - BusinessWeek

RSS Feed: Most Popular Slide Shows

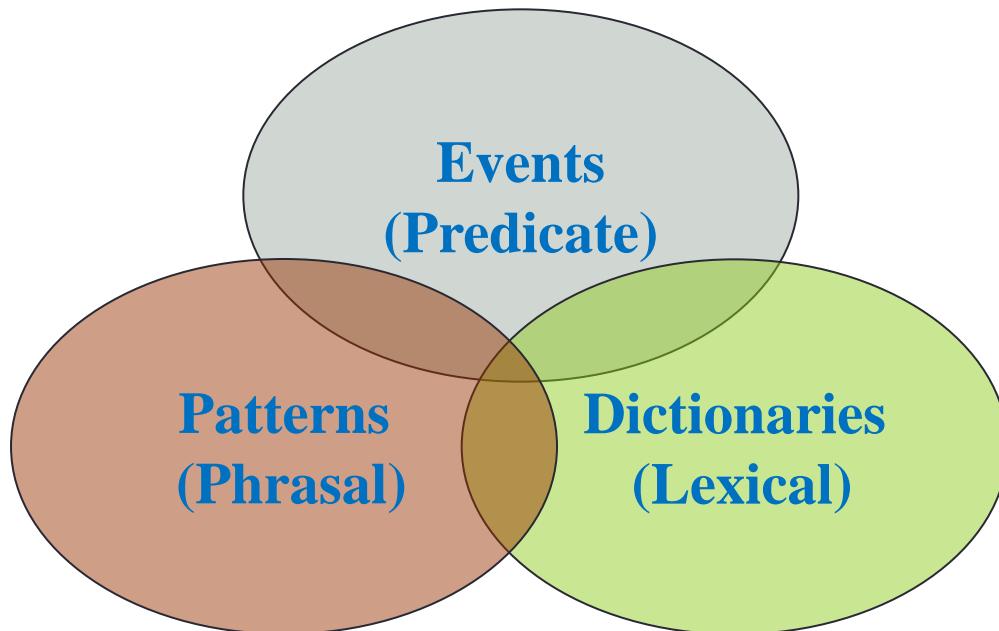
BW MALL - SPONSORED LINKS

[Tulane University Online Business Certificates](#) Earn a Master Certificate in Business Management or Marketing ONLINE in only 8 Weeks through top-ranked Tulane Freeman B-School. Learn co...

[Refinance Today!](#) Lock-In a 30 yr. Fixed Rate for as Low as 5.38%. Click Here

[24Hr Online Forex Trading](#) Trade EUR / GBP / USD / CHF / JPY & more. Up to \$200 bonus. Free Rates,

Hybrid Sentiment Analysis



- All levels are part of the same rulebook, and are therefore considered simultaneously by CaRE

Dictionary-based sentiment

- Started with available sentiment lexicons
 - Domain-specific and general
 - Improved by our content experts
- Examples
 - Modifiers: attractive, superior, inefficient, risky
 - Verbs: invents, advancing, failed, lost
 - Nouns: opportunity, success, weakness, crisis
 - Expressions: exceeding expectations, chapter 11
- Emphasis and reversal
 - successful, extremely successful,
far from successful

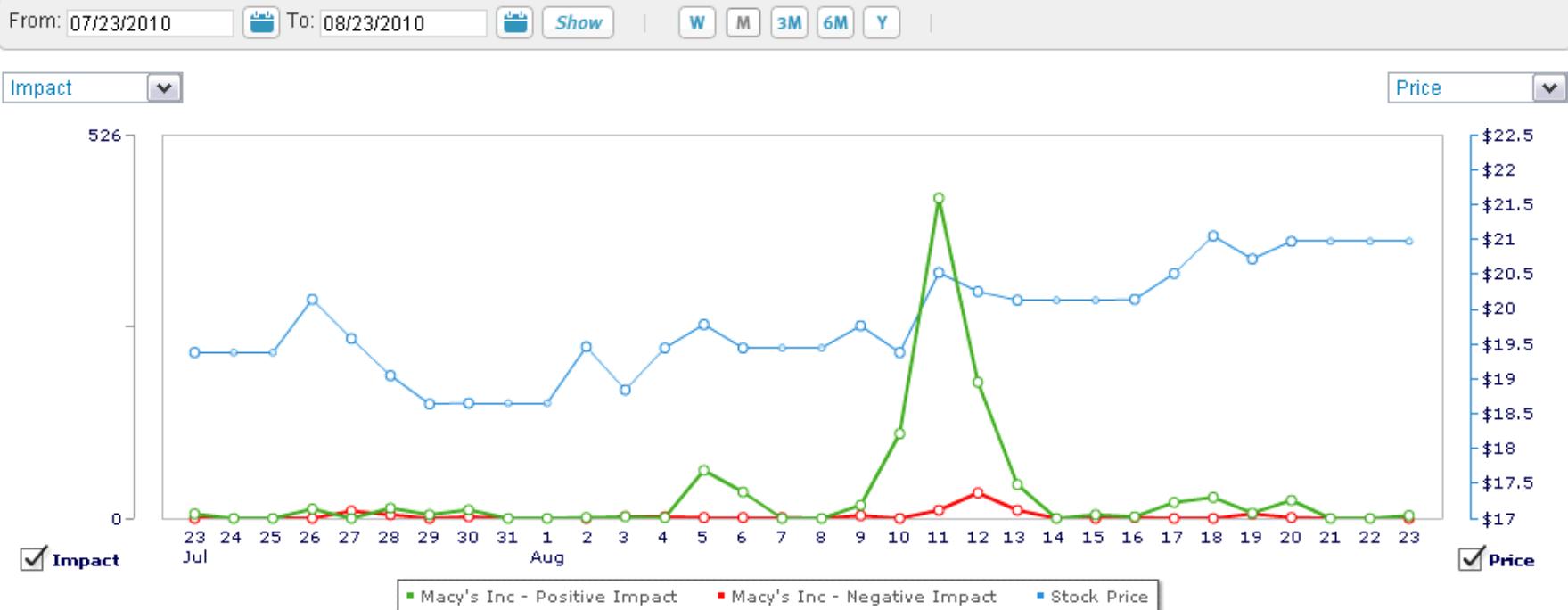
Event-Based Sentiment

- Product release/approval/recall, litigations, acquisitions, workforce change, analyst recommendations and many more
- Semantic role matters:
 - Google is being sued/is suing...
 - Need to address historical/speculative events
 - Google acquired YouTube **in 2006**
 - **What if Google buys Yahoo** and the software giant Microsoft remains a single company fighting for the power of the Internet?

Macy's

Macy's Inc (NYSE:M)

Add to Portfolio



JC Penny

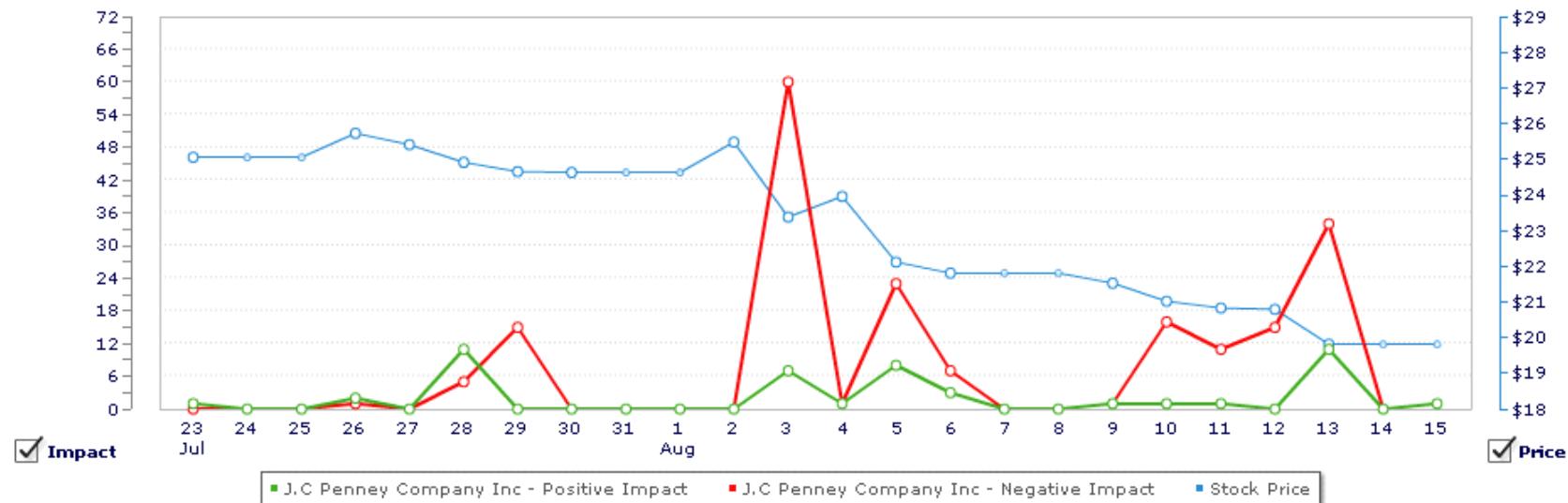
J.C Penney Company Inc (NYSE:JCP)

Add to Portfolio

From: 07/23/2010 To: 08/15/2010 Show | W M 3M 6M Y |

Impact

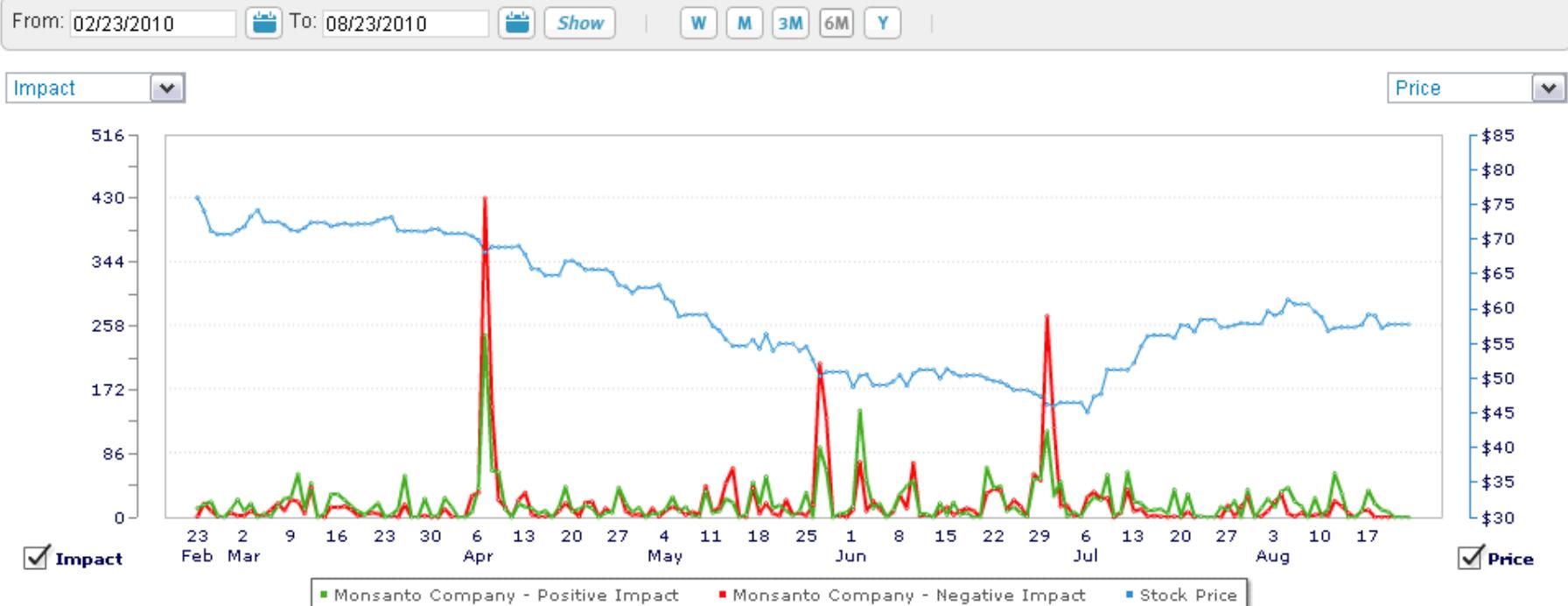
Price



Monsanto

Monsanto Company (NYSE:MON)

[★ Add to Portfolio](#)



Goldman Sachs

Goldman Sachs Group Inc (NYSE:GS)

Add to Portfolio

From: 07/23/2010

To: 08/23/2010

Show

W

M

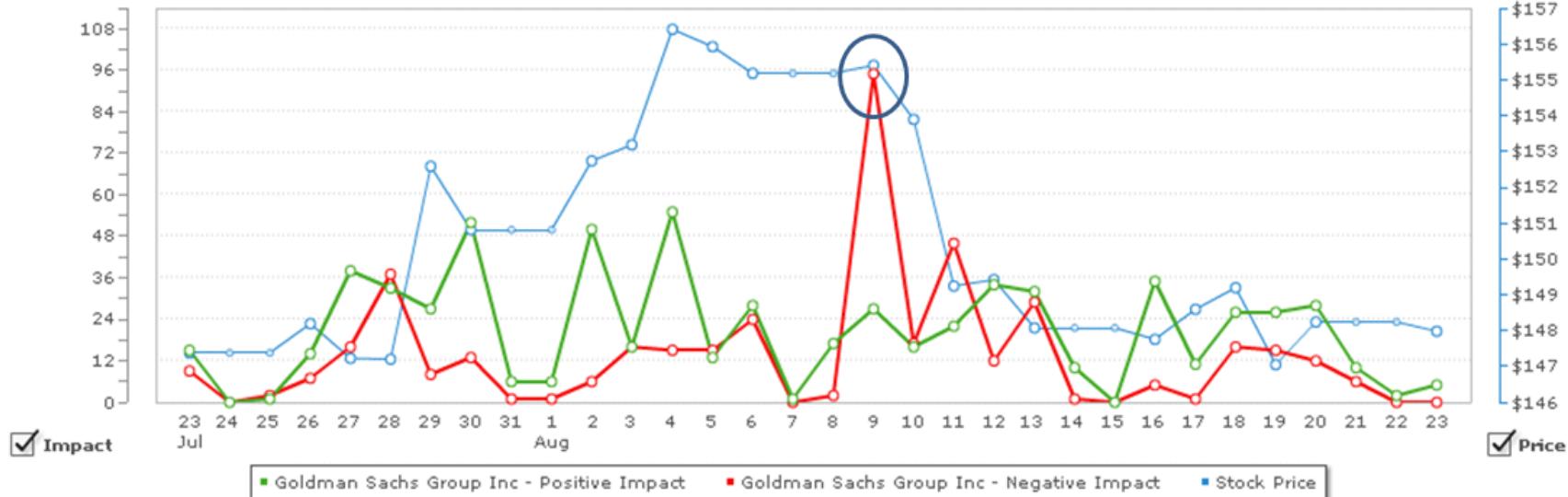
3M

6M

Y

Impact

Price



SNTA 3 Months

Synta Pharmaceuticals Corp (NASDAQGM:SNTA)

 Add to Portfolio

From: 09/18/2010

To: 12/18/2010

Show

W

M

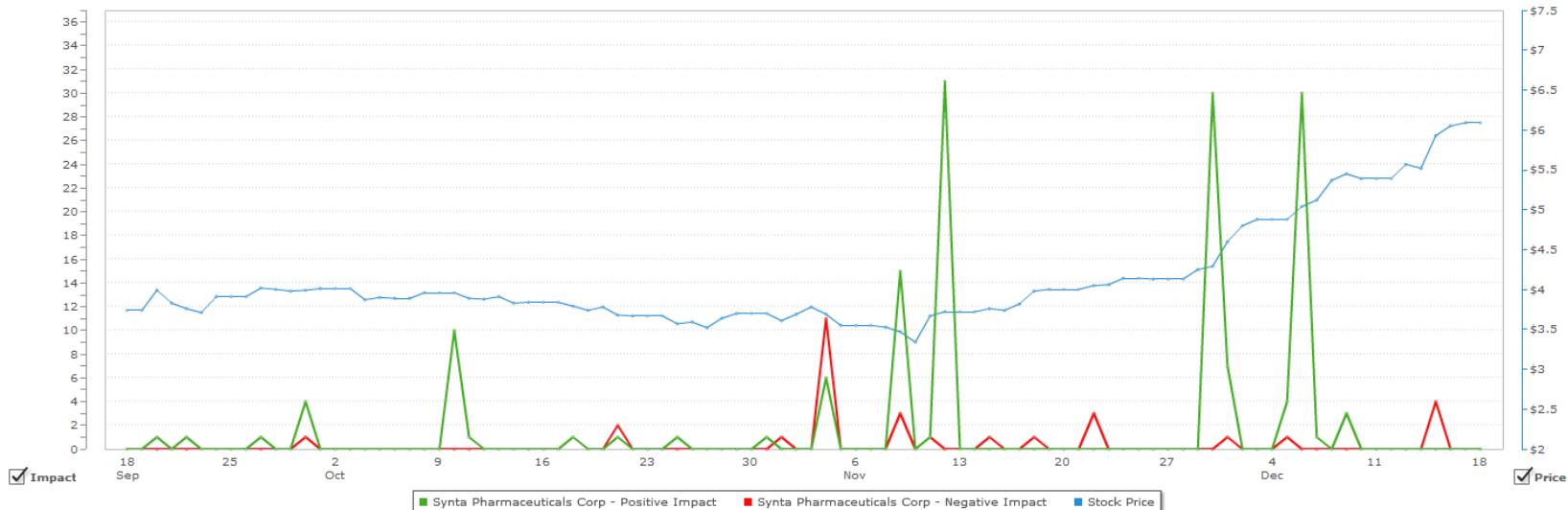
3M

6M

Y

Impact

Price

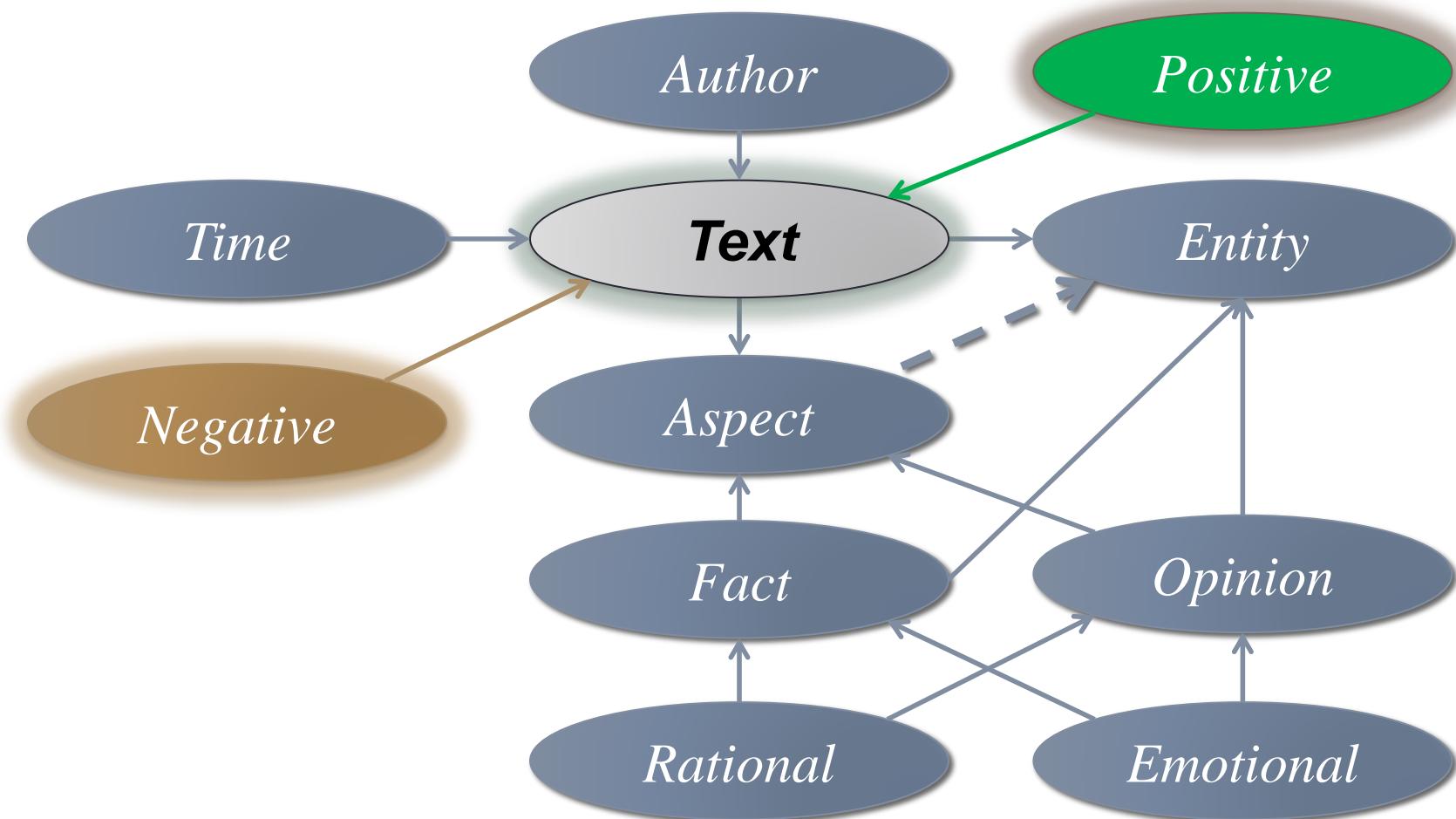


SENTIMENT ANALYSIS OF FINANCIAL TEXT-BASED MESSAGES IN MICROBLOGS

Why do we care about Microblogs' messages?

- Web 2.0 brought a significant change in the circles of human interactions.
 - People around the world are closer than ever before.
 - Nowadays, social media influences the way modern man makes decisions.
- Online investment message boards, such as Yahoo! Finance, Raging Bull and StockTwits:
 - Recently become very popular.
 - Allow investors to share trading ideas, advice and opinions on public companies.
 - Considered by many investors as highly valuable sources for making their trading decisions.
 - Can open an atypical window to typical investors' explicit thoughts, doubts as well as other opinions.

The complexity of Sentiment Analysis



Challenges related to sentiment analysis in general

- Sentiment words are domain-dependent (“The camera sucks” vs. “This vacuum cleaner sucks”).
- Sentiment words are aspect-dependent (“The weight of the computer is low ” vs. “The memory of the computer is low”).
- Sentiment modifiers (“highly”, “incredible”, “huge”)
- Emphasis modifiers (“mostly” and “quite”)
- Opposite modifiers such as “far from”
- Sentiment shifters (negation words such as “not”)
- Anaphora resolution
- Phrases (“reducing losses”) vs. words
- Sarcasm

Challenges related to sentiment analysis of Microblogs' messages

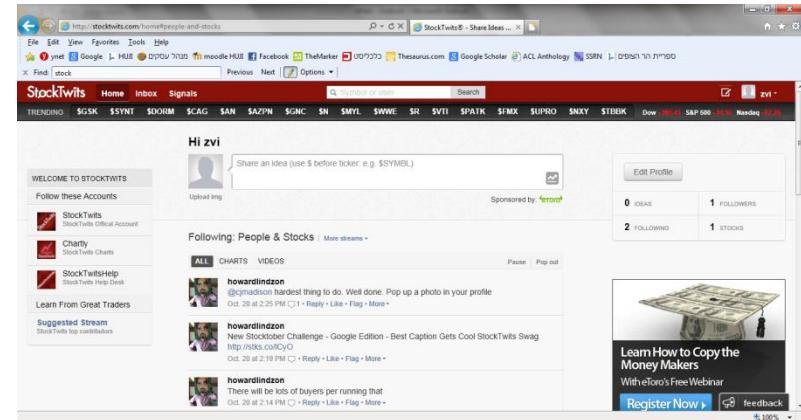
- Unique (Texting) style of language and ,often, sarcasm
- Textual structure is more diverse and complex than what is found in news articles.
- Use of short sentences, often omitting conjunctions, pronouns, etc.
- Conventional parsers and POS taggers perform poorly on Microblogs messages.

Research questions

- [1] Can investors' sentiments on social media be accurately identified?
- [2] Can certain true investors' sentiments on social media be exploited to predict stock price movement?
- [3] Do investors' sentiments that appear on social media correlate with returns?
- [4] Are investors' sentiments that appear on social media affected by news and prices?
- [5] What is the lead-lag relation between sentiments appearing on social media, sentiment appears on news article, and stock prices?
- [6] Do particular investors' expectations correlate with stock movements?

The dataset

- StockTwits
 - Started as a filtering service over the Twitter platform
 - Shows all twits that contain at least one stock ticker in them (A '\$' sign must be before the ticker of the stock to signal it is a ticker)
- Dataset including:
 - 2,750,000 Twits
 - Nearly 3 year period
 - July 2009 - May 2012
 - On average 1,703 twits per day
- Dow Jones' news archive of 22 years
 - 1990- June 2012



Tools

- Open source tools for textual analysis (Mainly in Python)
- A generic sentiment analysis engine;
- NER engine (based on CRF);
- The Care 2.0 Information Extraction Language;
- The Visual Care platform for the unsupervised extraction of relations between entities; and
- TSS (The Stock Sonar) System for identifying business events in text.

Methodological approach

- Differentiating between:
 - Relevant news for an entity to non-relevant news;
 - Expectations and past performances; and
 - “New” news versus “Old” news.
- Using background information in the analysis
 - Aimed at generating the sentiment lexicon, e.g.:
 - General-purpose Sentiment Lexicon;
 - Sentiment Rating; and
 - Thesaurus and Linguistic Heuristics.
 - Aimed at having deeper contextual understanding of the meaning of the messages, e.g.:
 - Stock price:
 - *“\$Big morning move on \$MU”*
 - *“\$SINA watch \$57.50 which now is a resistance, \$59.21 that will open the door to \$62.48 before earning, if not we will break \$50”*
 - Stock trend:
 - *“\$ES F http://stks.co/1k3d 1hour chart - Gap to fill - \$SPX \$SPY”*

04 January 2011

- \$MU FLYING GUYS
- RT @justmegregory shorted \$MU here again... you're shorting the biggest winner of 2011 in day 2? lol
- \$mu could be the stock of the year. i am speculating they will be in more apple products

15 March 2011

- \$MU staying long
- \$MU..getting interesting again
- Keeping an eye on \$MU today
- \$MU buy under \$10
- buying more \$MU here, increasing my position on this discount
- \$mu at low

01 April 2011

- WSJ - Best 10 S&P Stocks for the First Quarter
<http://on.wsj.com/dMvORZ> \$TSO \$MRO \$NFLX \$JDSU \$MU among others \$\$



01 March 2011

- \$MU Short @ 11.43
- \$mu \$apl looks like MU is not in apple ipad 2....sell as it ran up into thisyear. i am speculating they will be in more apple products

01 April 2011

- \$mu sucks @Superronen ur dreaming \$mu going down wait when markets go down \$ mu going t 7\$mu at low

20 April 2011

- \$MU out @ 11.20 (+.67 x 10k shr), finally it popped \$MU
- http://chart.ly/fkkoucu Cont.'s strong. May find some resistance around the 11.27 area.

26 April 2011

- \$MU short @ 11.50
- \$MU add more short @11.64
- \$MU cover short @ 11.58 (even),re-visit again tom. for possible market sell off

Preliminary results

C:\VC1\SAScripts\ResultsOct01RB.htm C:\VC1\SAScripts\ResultsO... File Edit View Favorites Tools Help y net Google HUJI moodle HUJI Facebook TheMarker כלכיסט Thesaurus.com Google Scholar ACL Anthology SSRN ספרית הר הצופים

News

Oct 2, 2012

- Aug_02_2009.txt
- Aug_03_2009.txt
- Aug_05_2009.txt
- Aug_06_2009.txt
- Aug_07_2009.txt
- Aug_08_2009.txt
- Aug_11_2009.txt
- Aug_12_2009.txt
- Aug_13_2009.txt
- Aug_14_2009.txt
- Aug_17_2009.txt
- Aug_18_2009.txt
- Aug_19_2009.txt
- Aug_20_2009.txt
- Aug_21_2009.txt
- Aug_22_2009.txt
- Aug_23_2009.txt
- Aug_24_2009.txt

Entity	Events	Patterns	Generic	Total Score
Ross Stores Inc. (ROST)	0/0	0/0	1/0	0.5
Avago Technologies Ltd (AVGO)	0/0	0/0	2/0	0.6666
British Pound Sterling (GBPJPY)	0/0	0/0	1/3	-0.4
PowerShares QQQ Trust, Series 1 (ETF) (QQQ)	0/0	0/0	0/1	-0.5
BioCryst Pharmaceuticals, Inc. (BCRX)				
ViaSat, Inc. (VSAT)				
Equity Financial Holdings Inc (GHC)	0/0	0/0	1/0	0.5
Reliance Steel & Aluminum (RS)	0/0	0/0	1/0	0.5
Crown Castle International Corp. (CCI)	0/0	0/0	1/1	0
Empire Resorts, Inc. (NYNY)				
iPath S&P 500 VIX Short Term Futures TM ETN (VXX)	0/0	0/0	1/2	-0.25
Analog Devices Inc (ADI)				
Trina Solar Limited (ADR) (TSL)	0/0	0/0	0/1	-0.5
British Pound Sterling (GBPUSD)	0/0	0/0	0/1	-0.5
SYNNEX Corporation (SNX)	0/0	0/0	0/1	-0.5
PDL BioPharma Inc. (PDLI)	0/0	0/0	1/0	0.5
Financial Select Sector SPDR (ETF) (XLF)				
Open Text Corporation (USA) (OTEX)	0/0	0/0	0/1	-0.5
Dell Inc. (DELL)				
Mercadolibre Inc (MEU)				
Ciena Corporation (CIEN)	0/0	0/0	1/0	0.5
Ann Inc (ANJN)				

News: Oct 2, 2012 Aug_19_2009.txt

is there a way to change the color of the bars in this app? having a hard time reading parts of it in default. leave the at tv integration
\$CCI interesting looking chart above base and entryy pullbacks. One to watch if markets show off weakness.
@bill which parts are you having a hard time reading?

SRS showing a possible topping pattern, but technically still at support.

lot's of bullish looking charts but \$JCLF isn't one of them
\$S highlighted some insurance companies in a recent article and they have been acting well. http://bit.ly/BDeo
\$Womex T2104 (stocks over 40m) still a robust 81.70% - tells me 2 things, market is still strong, and will soon be downside
\$TSL has been trying to fight the sector weakness, but testing on the edge of a small cliff here
\$GBJPY short at 1.55.00
\$GBJPY \$GBPUSD Perhaps there is an expectation that there will be bad news from the BOE policy minutes
RT. about \$GBJPY David Cameron releases British Default Warning, call is old story but it does matter http://bit.ly/vM8B
\$GBJPY \$GBPUSD Bank Voted 6 - 3 For 50 Bln QE Intrinsic http://bit.ly/tVQd
\$GBJPY \$GBPUSD CBI Industrial Order Expectations in 27 minutes. Be alert!
\$GBJPY covered all short positions for (1.55 x 2) = 310 pip gain
\$TSL was one of targets today, but gagging under my trigger
Works like a charm.
is there a way we can integrate the network?
@min Right now we have RSS integration but yes there is plan to integrate the network
@remmabatk Blocktraderide keep log file, the time stamp
how do i add a Twitter account so i can also follow my Twitter stream in here?
@bill Hi Bill go to the settings link in the top right. Then click the Add Twitter Button
Good morning all, getting ready to turn on the robot :-)

Precision	0.821192053
Recall	0.716763006
F1	0.765432099

Examples of correctly identified Twits

\$CSTR getting killed today

\$ININ also holding its breakout very well.

\$CISG nice volume surge and move today.

\$NOK still treading water, but hasn't broken yet. should be a decent short on a break of the past months range.

Experiencing pain in \$TCO Long from \$32.14

i am adding to shorts of \$xrt \$cof and now added \$axp. easy stop points above ...with \$etfc rising, we are near a top

\$GFIG still setting up nicely on daily chart. looks good if it can get over 7.30-7.40ish.

\$aig short 43.45

Once we make it past 1900 \$SI F - which will be very sudden when it happens....shortly afterwards we will hit new bull market highs. \$SLV

In \$VSH at 7.61 and out at 7.65 -I liked its slope up and the ratio of its bid-to-ask quantities

Examples of correctly identified Twits

Another Eastman Kodak Director Departs, Fleeing **Sinking Ship** <http://stks.co/1jUc> **\$EK**

Prices are LOCKED on **our 10 buy-and-holds** for 2012 - **\$AA \$ARCO \$CAT \$COF \$FDX \$HSY \$MAKO \$MSFT \$STD \$TKC** <http://stks.co/1jRN> #beststocks

\$SMH <http://stks.co/1jWX> actually **\$SOX semiconductor index gave up** 1% for the week closing under both moving averages again

RT @Glibonacci **Long \$CTRP** @ 24.13, stop under **recent low**. Looks primed for a reversal.

Yearly. Severe p/b to the 50% retracement. Look for a spot there for **a bounce** near the 22 level. **\$CTRP** <http://chart.ly/vwwvxh8>

Yearly. **3 year bounce** supported by a nice ^ move in volume. **\$VOD** <http://chart.ly/47493ys> **\$BKX** <http://stks.co/1jX8> banking index **dropped Tues +Wed , rallied back on Thursday with some consolidation on Friday.**

@1Investor After Jan, Apr, Jul earnings rpts, I am betting (BIG) that **\$BIDU** is above that **Shooting Star's top!** Happy/Prosperous 2012 TO YOU!

\$TDG <http://stks.co/1jSo> Daily chart shows **possible buy point around the \$100 price level** which is where my alert is set.

New Post: "Small Cap MLPs **Beware: Short Seller Takes Down STON**" <http://stks.co/1jkr> **\$STON \$GMCR**

Examples of correctly identified Twits

\$WG <http://stks.co/1jg3> Long setup: RSI positive divergence, triggers over 3.80

\$HDY Hammer candle in new trend, after sharp downtrend = Bullish, should see a nice pop tomorrow. We shall see comming morning.

\$HDY if overall market action is positive, we should HDY either A.Hit 2.60 or B. stay stagnant with a range of 2-3 cents from open. Prob A

\$HDY A/D turning upwards...hammer cand, stochastics traveling upwards as well..can see \$3 being touched in the next 12 Trading Days.

\$HDY <http://stks.co/1jlW> This one is set to go. Tax loss. Thinking it will gap up Tuesday. Don't chase. I'm holding from 2.26

@BumbleSpoon concur, A/D, MFI turning up, i think \$HDY can have a soild run to 2.80-3\$ lower bands were tag its a positive Volatility play

\$LINTA <http://stks.co/1jnA> Long setup. Looks god over 16.50.

\$FLT <http://stks.co/1jcl> Nice spike in weekly and daily volume in December, looks ready to clear 30 bucks per share

\$NM <http://stks.co/1jbh> Double bottom followed by a symmetrical triangle.

Examples of current precision errors

I like the action in **\$BWLD** here too. a few points under all time highs, but also some stiff resistance.

@stevenplace also, couple days ago, you like **\$RIMM** short below 78, still feel the same?

just ran through the non **\$RAD** drug store stocks and all look decent except for **\$OCR** - might be worth watching as a short if group pullsback

writing an article on airline stocks tonight, just found out they have an etf - **\$FAA** , looks pretty bullish but very thin trading

also watching two low volume oil stocks that I think could really get moving - **\$PDO** and **\$HKN**

My very first Wallstrip show... **\$AAPL** <http://bit.ly/1a0za5> have a great weekend. BIG NEWS TUESDAY for us stocktwiterers

Could **\$AAPL** crush TV? **\$MSFT** thinks so. --> <http://bit.ly/NNbQ3>

Examples of current recall errors

looks like **\$BBT** bought Colonial Bank, which is causing the spike in price and volume

\$BBT getting some volume on the upside

\$HURN has been persistently strong this morning. added to my watchlist for later today.

nothing on the mojo list this morning. sticking with **\$FAZ** and **\$TZA** for now.

\$GGC at 08:28:32 8/18/2009: Running up: +0.27 in less than one minute

\$BCRX getting a pop

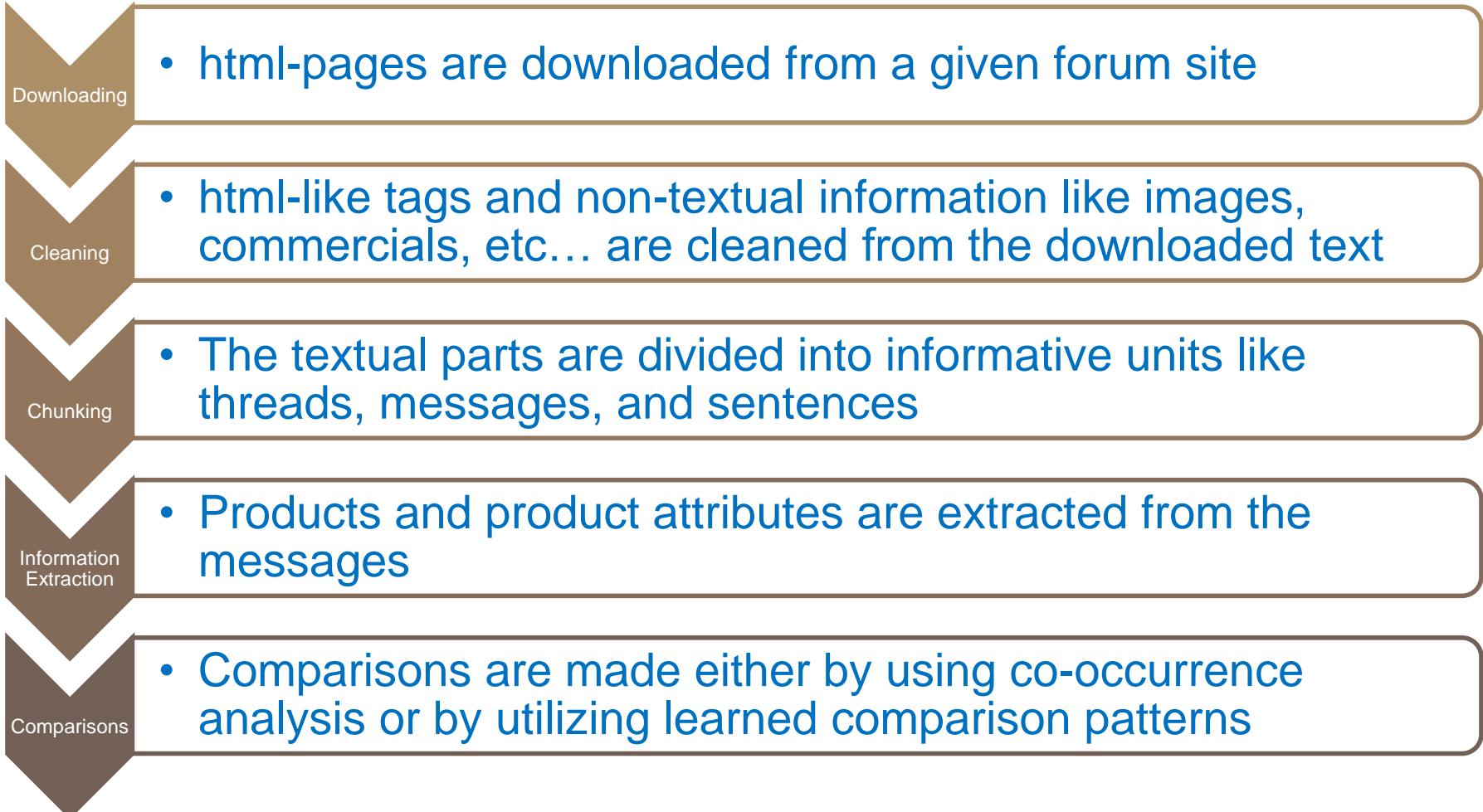
surge in **\$NYNY** as well. lots of these little stocks moving today.

\$ANN **\$ADI** and **\$CIEN** on my @tradeideas mojo filter today.



Mining Medical User Forums

The Text Mining Process



The Text Mining Process



We downloaded messages from 5 different consumer forums

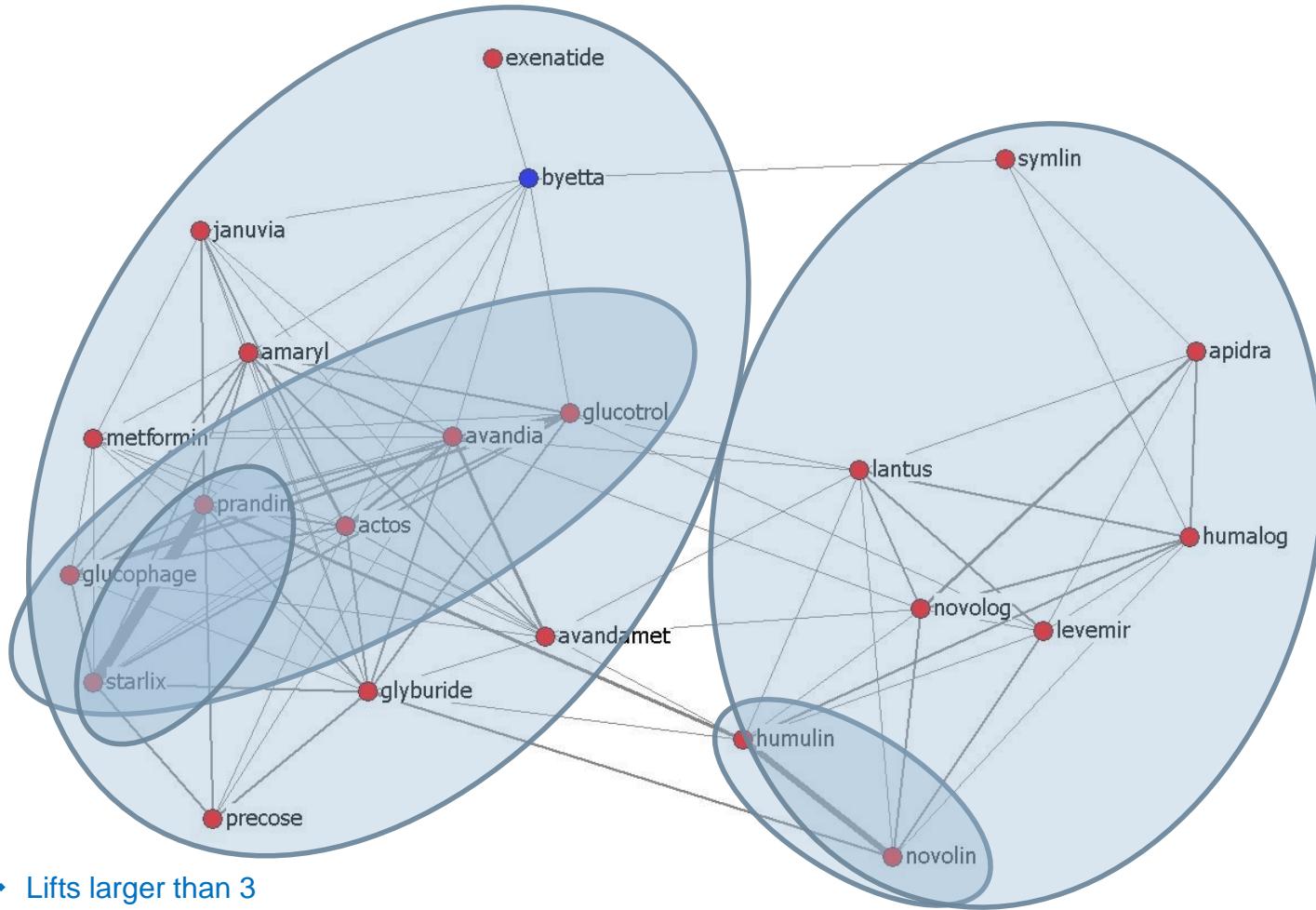
- diabetesforums.com
- healthboards.com
- forum.lowcarber.org
- diabetes.blog.com**
- diabetessdaily.com



** Messages in Diabetess.blog.com were focused mainly on Byetta

Drug Analysis

Drug Co-Occurrence - Spring Graph – Perceptual Map



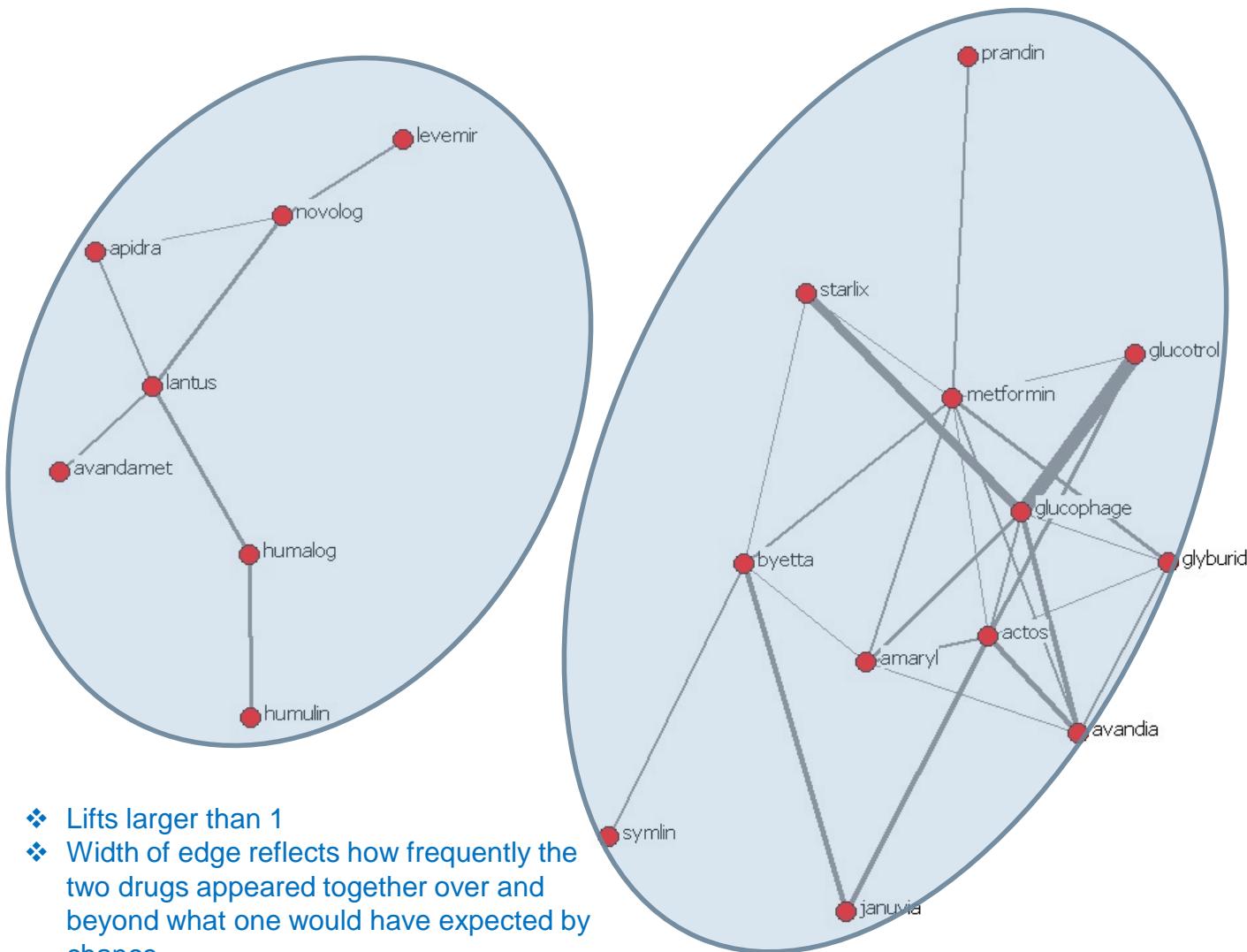
Several Pockets of drugs that were mentioned frequently together in a message were identified. Byetta was mentioned frequently with:

- Glucotrol
- Januvia
- Amaryl
- Actos
- Avandia
- Prandin
- Symlin

- ❖ Lifts larger than 3
- ❖ Width of edge reflects how frequently the two drugs appeared together over and beyond what one would have expected by chance

Drug Usage Analysis

Drug Co-taking – Drugs mentioned as “Taken Together”



There are two main clusters of drugs that are mentioned as “taken together”

Byetta was mentioned as “taken together” with:

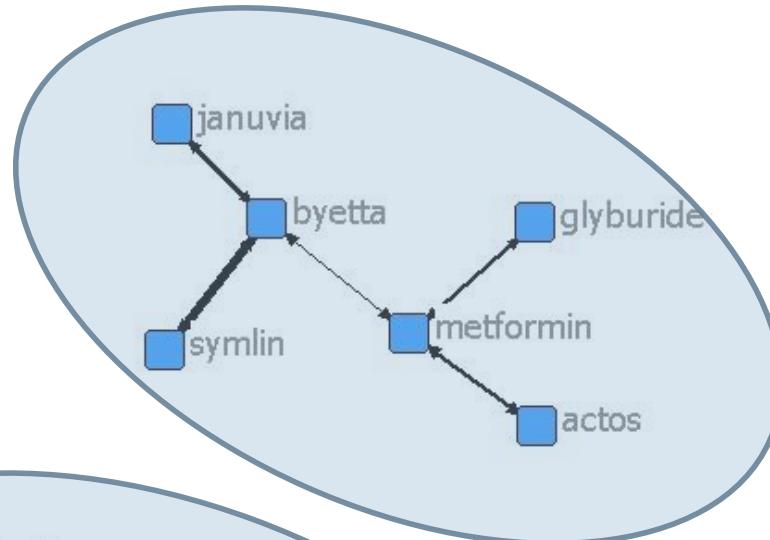
- Januvia
- Symlin
- Metformin
- Amaryl
- Starlix

Pairs of drugs that are taken frequently together include:

- Glucotrol--Glucophage
- Glucophage--Stralix
- Byetta--Januvia
- Avandia--Actos
- Glucophage--Avandia

Drug Usage Analysis

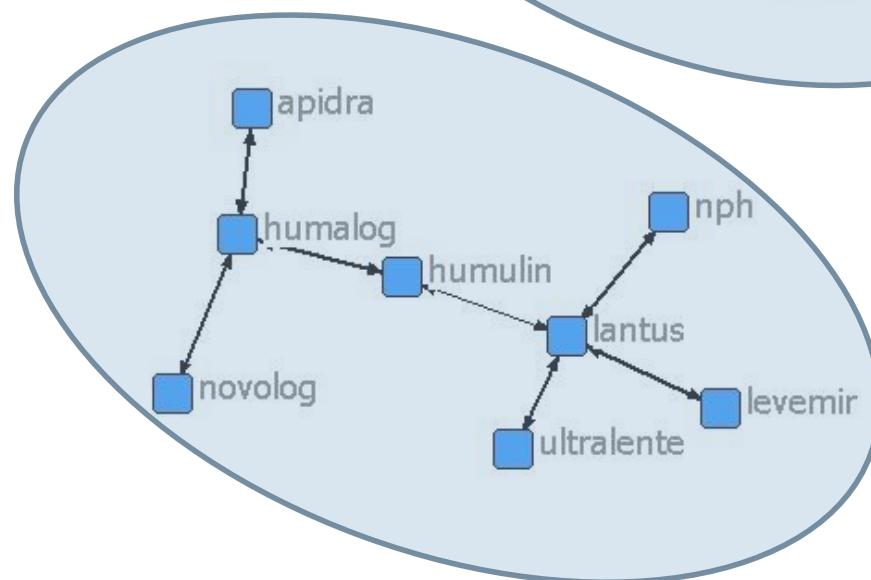
Drug Switching – Drugs mentioned as “Switched” to and from



There are two main clusters of diabetes drugs within which consumers mentioned frequently that they “switched” from one drug to another

Byetta was mentioned as “switched” to and from:

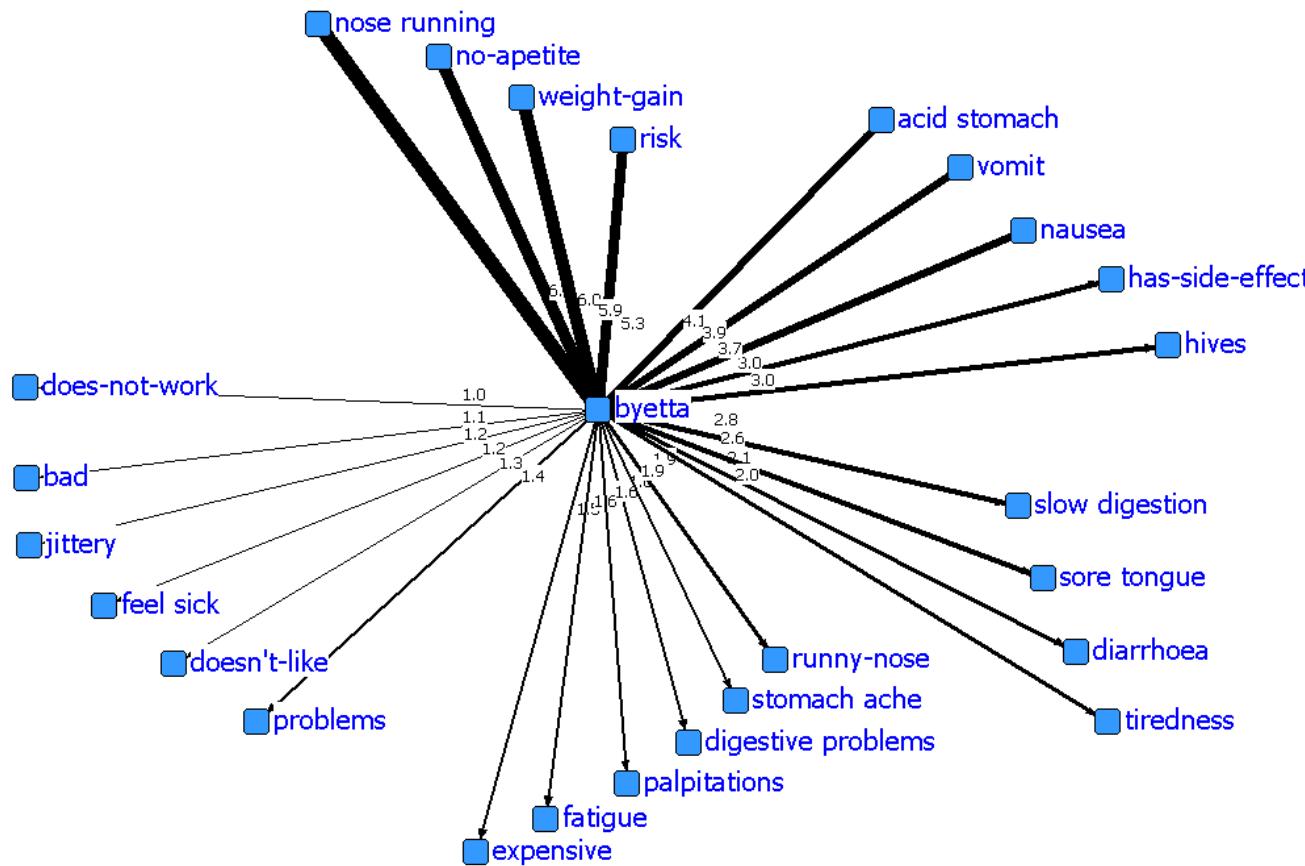
- Symlin
- Januvia
- Metformin



- ❖ Lifts larger than 1
- ❖ Width of edge reflects how frequently the two drugs appeared together over and beyond what one would have expected by chance

Drug Terms Analysis

Byetta - Side Effects Analysis

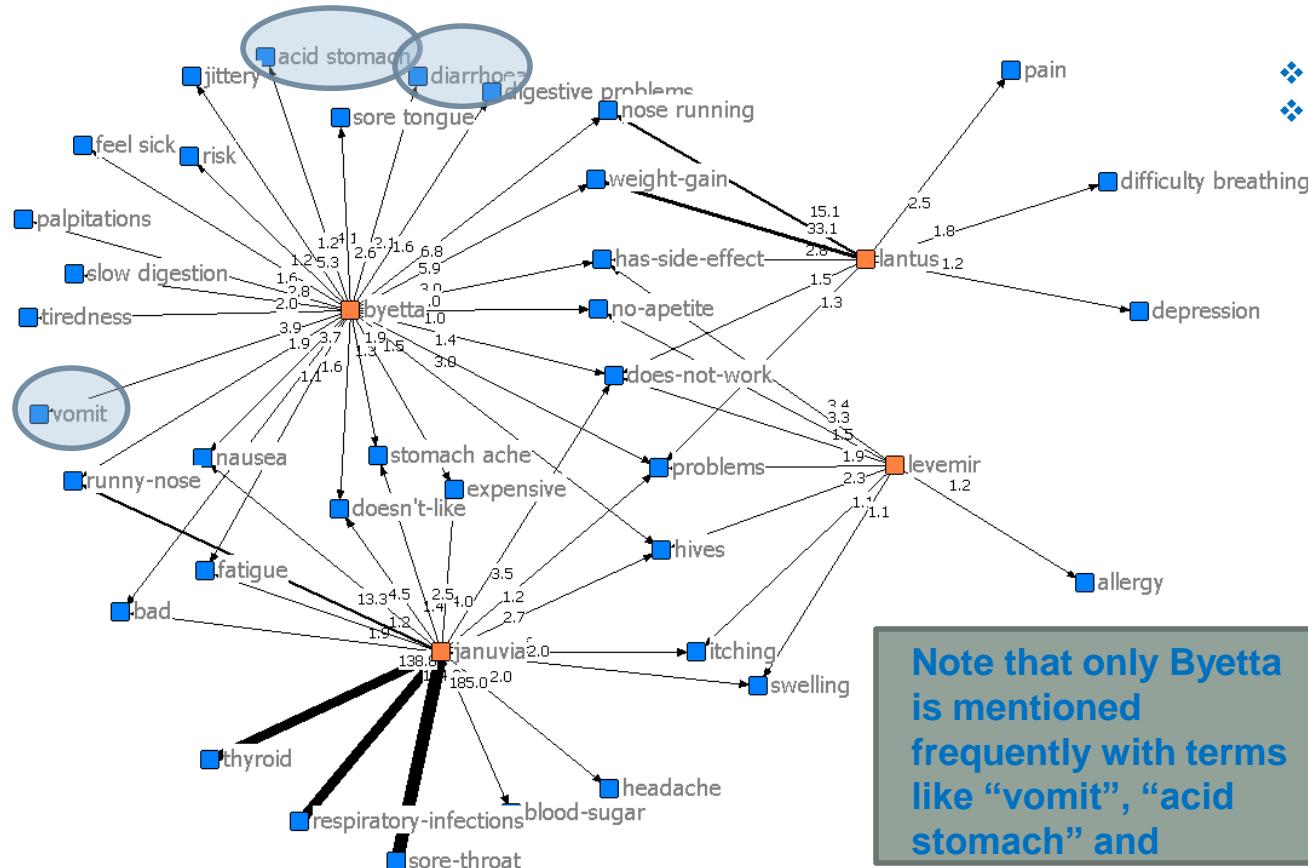


Byetta appeared much more than chance with the following side effects:

- “Nose running” or “runny nose”
- “No appetite”
- “Weight gain”
- “Acid stomach”
- “Vomit”
- “Nausea”
- “Hives”

Drug Terms Analysis

Drug Comparisons on Side Effects



Byetta shares with

Januvia the side effects:

- Runny nose
 - Nausea
 - Stomach ache
 - Hives

Byetta shares with Byetta shares with

Januvia the side effects: | eyemir the side effects:

- Runny nose
 - Nausea
 - No appetite
 - Hives

Note that only Byetta is mentioned frequently with terms like “vomit”, “acid stomach” and

“diarrhea”

Byetta shares with Lantus the side effects:

- Nose running
 - Weight gain

- ❖ Lifts larger than 1
- ❖ Width of edge reflects how frequently the two drugs appeared together over time and beyond what one would have expected by chance

The main side effects discussed with Januvia:

- Thyroid
 - Respiratory infections
 - Sore throat

The main side effects discussed with Levemir :

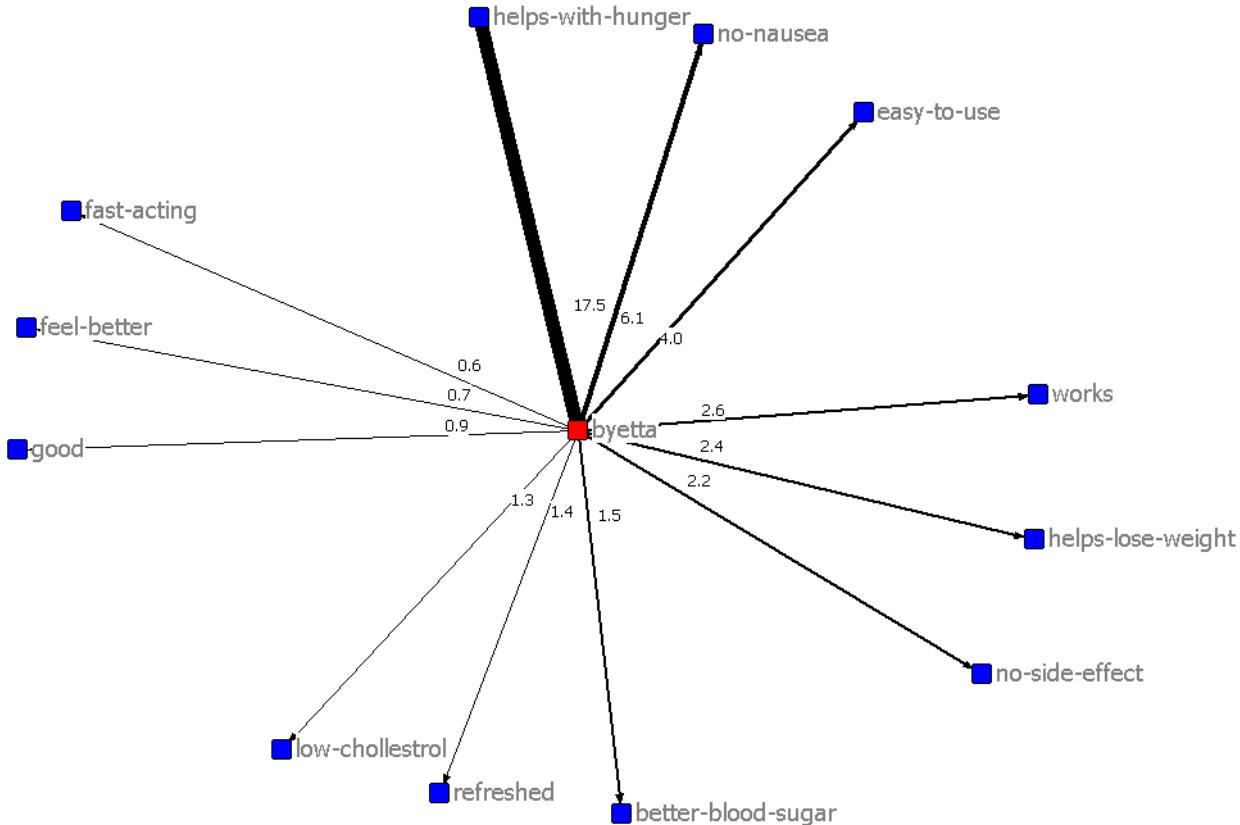
- No appetite
 - Hives

The main side effects discussed with Lantus:

- Weight gain
 - Nose running
 - Pain

Drug Terms Analysis

Byetta – Positive Sentiments

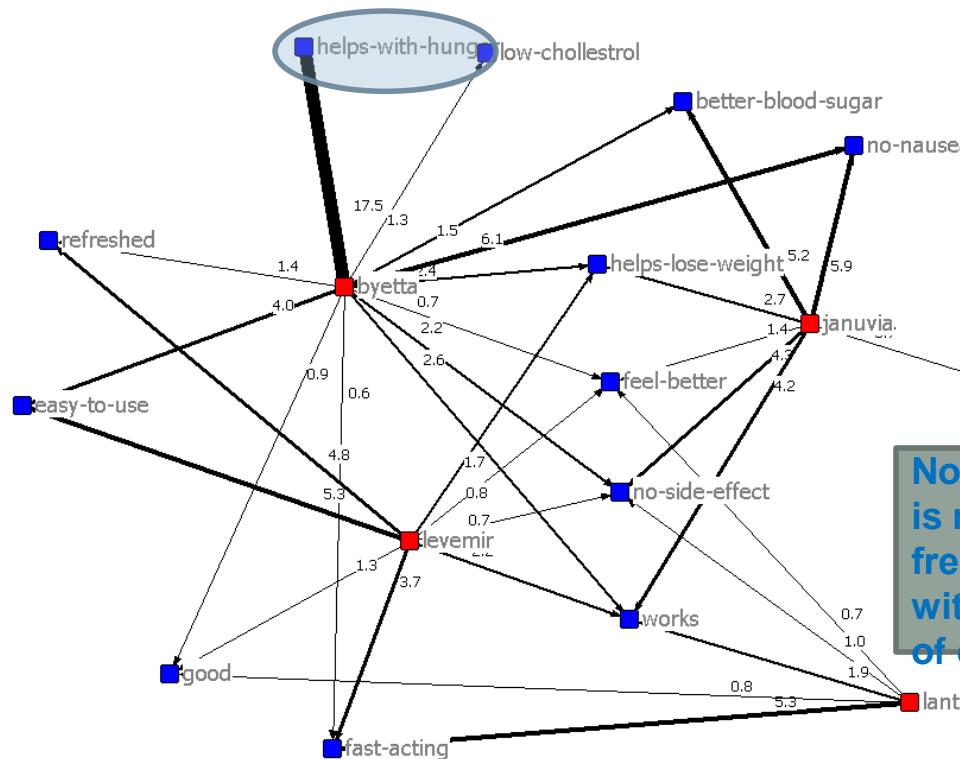


Byetta appeared much more than chance ($\text{lift} > 2$) with the following positive sentiments:

- “Helps with hunger”
- “No nausea”
- “Easy to use”
- “Works”
- “Helps losing weight”
- “No side effects”

Drug Terms Analysis

Drug Comparisons on Positive Sentiments



Byetta shares with Januvia:

- “Better blood sugar”
- “No nausea”
- “Helps lose weight”
- “No side effects”
- “Works”

Byetta shares with Levemir:

- “Easy to use”
- “Helps lose weight”
- “No side effects”
- “Works”

Byetta shares with Lantus:

- “Easy to use”
- “No side effects”
- “Works”

- ❖ Lifts larger than 0.5
- ❖ Width of edge reflects how frequently the two drugs appeared together over and beyond what one would have expected by chance

The main positive sentiments discussed with Januvia:

- “No nausea”
- “Better blood sugar”
- “Works”
- “No side effects”

The main positive sentiments discussed with Levemir:

- “Easy to use”
- “Fast acting”

The main positive sentiments discussed with Lantus:

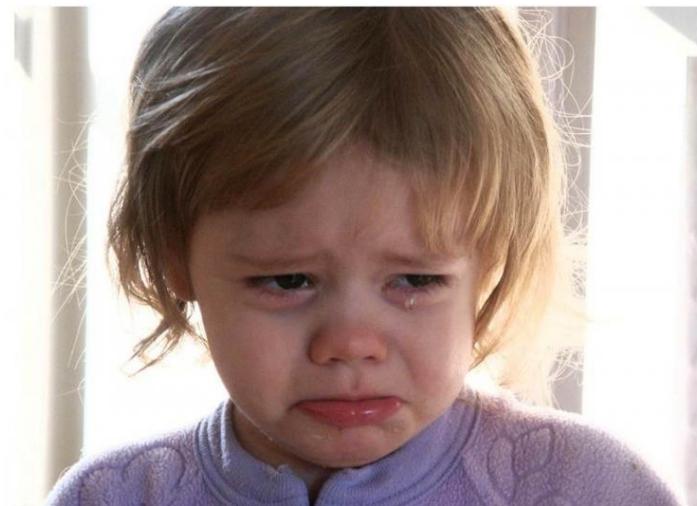
- “Fast acting”
- “Works”

Sentiment in Social Media

- The words people use on social media such as Twitter, Facebook and Google are a rich, if imperfect, source of information about their personality and psychological state.
- Sentiment analysis of tweets and blogs has been used to chart variation in happiness over time and space.
- We show how variation in Facebook word use with age, sex, and personality suggests new hypotheses for psychological research.

Our goal: Measure the world's well-being using social media

- Subjective well-being matters
 - Happy people with better relationships live longer
 - We want *better* as well as *happier* lives
- The words people use reveal a lot about them
 - including personality and happiness



First step: measure personality

PERMA measures well-being

- Positive emotion
- Engagement
- Relationships
- Meaning
- Achievement



OCEAN measures personality

- Openness
- Conscientiousness
- Extroversion
- Agreeableness
- Neuroticism



New tools for psychology

- Typical psychology study
 - 50 undergraduates

Hypothesis testing

- Gallup Opinion Poll
 - 1 million surveys

Monitoring

- Social media



- 60 million status updates/day



- 300 million Tweets/day



- 3 billion searches/day

Hypothesis generation

open vocabulary data-driven exploration

Basic Sentiment Analysis

facebook    

Search 

Johannes Eichstaedt Edit My Profile

News Feed

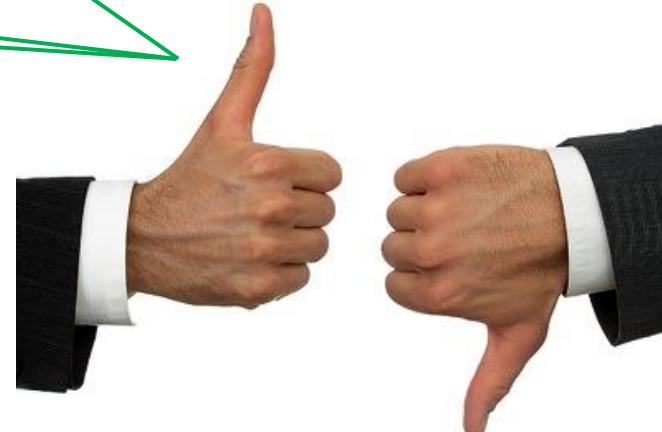
Share:  Status  Question  Photo  Link  Video

What's on your mind?

Steph Ramones  Had a great 5 min today, personal best post injury. Hopefully this will get me a PR for a 10k (yay MAPP mini marathon) next week.
5 hours ago · Like · Comment 
 View all 7 comments
Carin Zinter @Leona, I'm walking in solidarity with my un-trained peeps (feel the love Jamie!). In reality, I've been more focused on my martial arts training the past few months -- hoping to test for my brown belt in August.
3 hours ago · Like
Leona Bratton @Leona Glad to hear you are not injured, you are a picture of empathic relationships in action, and concentrate on the brown belt! :)
2 hours ago · Like

Carin Rockind Chicago Tribune reports about how I used Positive Psychology to change my own attitude and life.

 **Happiness, meet science** bit.ly
Carin Rockind had a condo, a boat, a great job in corporate marketing and dozens of friends and business contacts, all based in Cleveland.
 Yesterday at 10:42am · Like · Comment · Share



Do you like the product?

Sentiment analysis in marketing

- What are reviewers and bloggers saying about my product -
 - or me?
 - What do people like or dislike about it?
 - What products do they compare it against?
 - On what dimensions?
 - How are their opinions or priorities changing?



Dynamic, large scale market research

Psychologists also use word lists

- Correlate percentage of words from a list
 - usually positive vs. negative affect
- with some outcome
 - Find (e.g.) that nuns using positive affect words live longer

LIWC dictionary sample

- accept
- accepta*
- accepts
- advantag*
- agreeing
-
- happy
- ...
- merry
-
- heartfelt
- heaven*
- heh*
- helper*
- helps
- hilarious
- hoho*
- honest*
- honour*
- hoped
- hopefully
- hopes
- terrific*
- thank
- thanked
- thanks
- toleran*
- treasur*
- treat
- trueness
- truer
- truest
- truly
- trust*
- values
- valuing

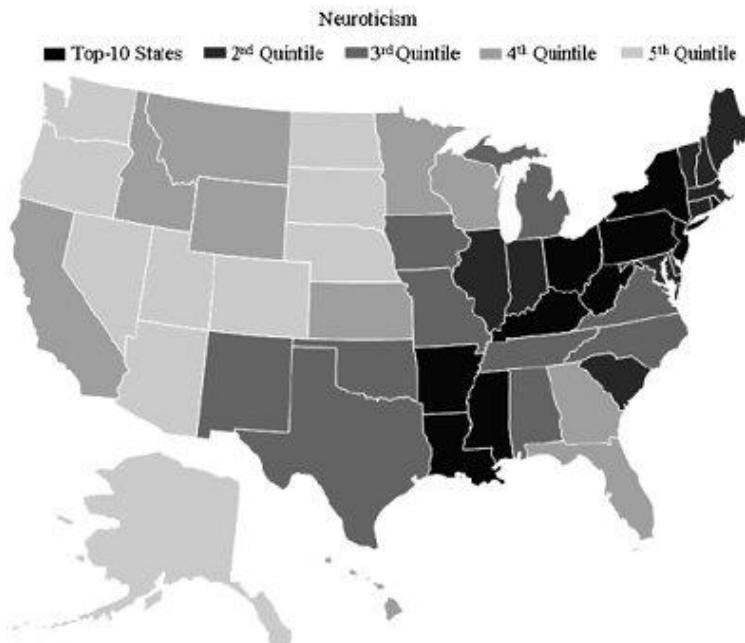
Pennebaker

We want to
Go beyond positive and
negative affect

Get a picture of people's
personalities, relationships, and
concerns

Facebook Individual Data

- 100,000 individuals who took Big Five Personality test
 - And their Facebook posts



Neuroticism by state:
Rentfrow, Gosling, and Potter

Kosinski, M. &
Stillwell, D.J.
(2011)

GENDER

Female

Male

PERSONALITY: BIG 5

Less Agreeableness

A word cloud centered around the word "fucking" in large red font. Other prominent words include "bitches", "shit", "assfuck", "pissed", "hell", and "dumbass". The words are colored in shades of red, blue, purple, and grey, and are surrounded by numerous other curse words and slang terms.

More Agreeableness

A word cloud centered around the theme of agreeableness, featuring words like praise, blessed, amazing, wonderful, and prayer.

The word cloud includes the following words:

- happiness
- good_day
- god_bless
- miss_you
- savior
- super_excited
- birthday_wishes
- strength
- practice
- encouragement
- to_see_my
- in_jesus
- exciting
- thanks
- praise
- happy_easter
- ready_for
- the_beach
- grace
- day_today
- study
- the Lord
- grateful_friends_and_family
- great_weekend
- hope_everyone
- of_july
- church
- a_great
- having_a_great
- pastor
- love_you_guys
- blessed
- isaiah
- spirit
- joy
- god_has
- love_you_all
- a_blast
- new_year_!
- praying
- spring
- thank_you
- merry_christmas
- proverbs
- beautiful_day
- excited
- faithful
- haha
- concern
- thankful
- amazing
- god's
- thanksgiving
- pray_for
- goodness
- forward_to
- be_there
- this_morning
- game_tonight
- every_moment
- !_can't_wait
- blessings
- wonderful
- in_christ :)
- you_smile
- homework
- anAwesome
- pumped
- camp
- relaxing
- corinthians
- prayer
- kindness
- tomorrow
- much_fun_!
- your_heart
- summer
- lots_of_loving
- worship
- so_glad
- romans
- trust_in
- snowboarding
- last_night_!

Less Consciousness

More Conscientiousness

A word cloud centered around the theme of conscientiousness, featuring words like relax, blessings, vacation, thankful, great, day, ready, for, weekend, work, wonderful, long, day, workout, busy, success, family, and forward.

The word cloud includes the following words:

- much-needed
- relax
- dinner_with
- night_with another_day
- so_glad
- finals
- work_tomorrow
- safe_and
- hard_work
- excited
- blessings
- more_days
- done_with
- thank
- care_of opportunity
- this_week
- lunch_with
- vacation
- thankful
- day_off
- training
- exhausted
- appreciate
- hubby
- headed_to
- pray
- blessed
- the_gym
- great
- day
- ahead
- much_to_do
- studiing_for
- obstacles
- challenges
- thankng
- made_it
- days_until
- study
- my_job
- loving
- working
- sunday
- counting
- the_lord
- ready_for
- relaxing
- semester
- bless
- last_day
- tgif
- an_amazing
- the_weather
- everyone_for
- prayers
- much_fun
- thanks_final
- will_be
- enjoyed
- leadership
- super_excited
- :-)
- weekend
- to_work
- ga_blast
- classes
- thanksgiving
- successful
- graduation
- on_my_way
- fam
- heading
- god_has
- wonderful
- long
- day
- workout
- degree
- strength
- time_with
- birthday_wishes_!
- wait_to
- on_my_way
- patient
- college
- congrats
- beautiful_day
- enjoying
- this_morning
- mother's_day
- here_we_come
- great_friends
- spending_time
- busy
- success
- family
- forward

Less Extraversion (Introversion)

More Extraversion

A word cloud centered around extraversion, featuring words like party, love, baby, and chillin.

The word cloud includes the following words:

- chillsoo
- bout
- ya ;)
- ladies
- soooo
- doin
- last_night_!
- great_night
- sooooo
- weekend
- hit_me_up
- great_friends
- time_with
- whats
- much_fun
- tonight
- partying
- then_off
- jersey_shore
- ;))
- !_!_! a_blast
- great_night
- comin
- sunday
- great_friends
- great_night
- tanning
- its_gonna
- havin
- ur
- wait
- fam
- gotta
- party
- goin
- ill
- tryin
- haha
- ready
- gettin
- love
- you
- night
- with
- its
- loves
- holla
- jersey
- football
- sexy
- haters
- cuz
- beach
- thinkin
- chillin
- baby
- my_life
- an_amazing
- miss
- pumped
- didnt
- workin
- excited
- big
- to_see_my
- wanna
- right_now_!
- babe
- chillin
- bestie
- thats
- im
- lovin
- girls
- out_with
- soooo
- here_we
- text_me
- lets
- girl
- friday
- excited
- big
- to_see_my
- wanna
- right_now_!
- call_me
- missin
- dance
- feelin
- gym
- bday
- the_best
- aunt
- boys
- wit
- pool
- love_u
- yall
- lookin
- lil
- on_my_way
- we_come
- blessed
- guys
- beautiful

Less Neuroticism

More Neuroticism

A word cloud centered around neurotic and depressive language. The most prominent words are "depression" (grey), "fuck" (red), "sick" (blue), "of" (grey), and "fucking" (red). Other significant words include "depressed" (grey), "lonely" (blue), "shit" (red), "angry" (blue), "bored" (blue), "hates" (blue), "crying" (blue), "alone" (blue), "kill" (blue), "xdanymore" (red), "stupid" (red), "broken" (blue), "tears" (blue), "worried" (blue), "the_point" (blue), "twat" (blue), "nervous" (blue), "mood" (blue), "hurts" (blue), "internet" (blue), "actually" (blue), "scared" (blue), "leave_me" (blue), "won't" (blue), "fed_up" (blue), "annoying" (blue), "bother" (blue), "horrible" (blue), "headache" (blue), "care" (blue), "go_away" (blue), "depressing" (blue), "scream" (blue), "anxiety" (blue), "anime" (blue), "internet" (blue), "actually" (blue), "for_at_leastpathetic" (blue), "for_once" (blue), "mum" (blue), "panic" (blue), "why_do_I" (blue), "tired_of" (blue), "emo" (blue), "nightmare" (blue), "shitty" (blue), "as_your_status" (blue), "seem_to_I'd" (blue), "film" (blue), "cure" (blue), ""><.
The words are colored in shades of red, blue, and grey, with some words having multiple color variations. The size of each word indicates its frequency or importance in the context of the word cloud.

Less Openness to Experience

Ung ung walang tlga
cnt any_1 gud ur nd
nman im XXXX wat happy_new_year 2tym
u_r naman naman XXXXX every_1 cant_wait_till Ing lang
.._haha goin_to! sum_1 mum (:_1 nt araw alot nyo tau wag coz
1st ?4? goin_to! sum_1 mum every_1 cant believe y salamat taong sarap nxt yan
dont:(cant_(believe y salamat .. :) na 2day wala
jst luv cant believe y salamat .. :) na 2day soo
nite nyt you_are_proud game_tonight 2day soo
xx miss_u had_a_good :@
hubby its_friday (:_) bt didnt wen hirap
x_x its_friday (:_) bt didnt wen hirap
x_x dnt buhay sana alam mong isang

More Openness to Experience

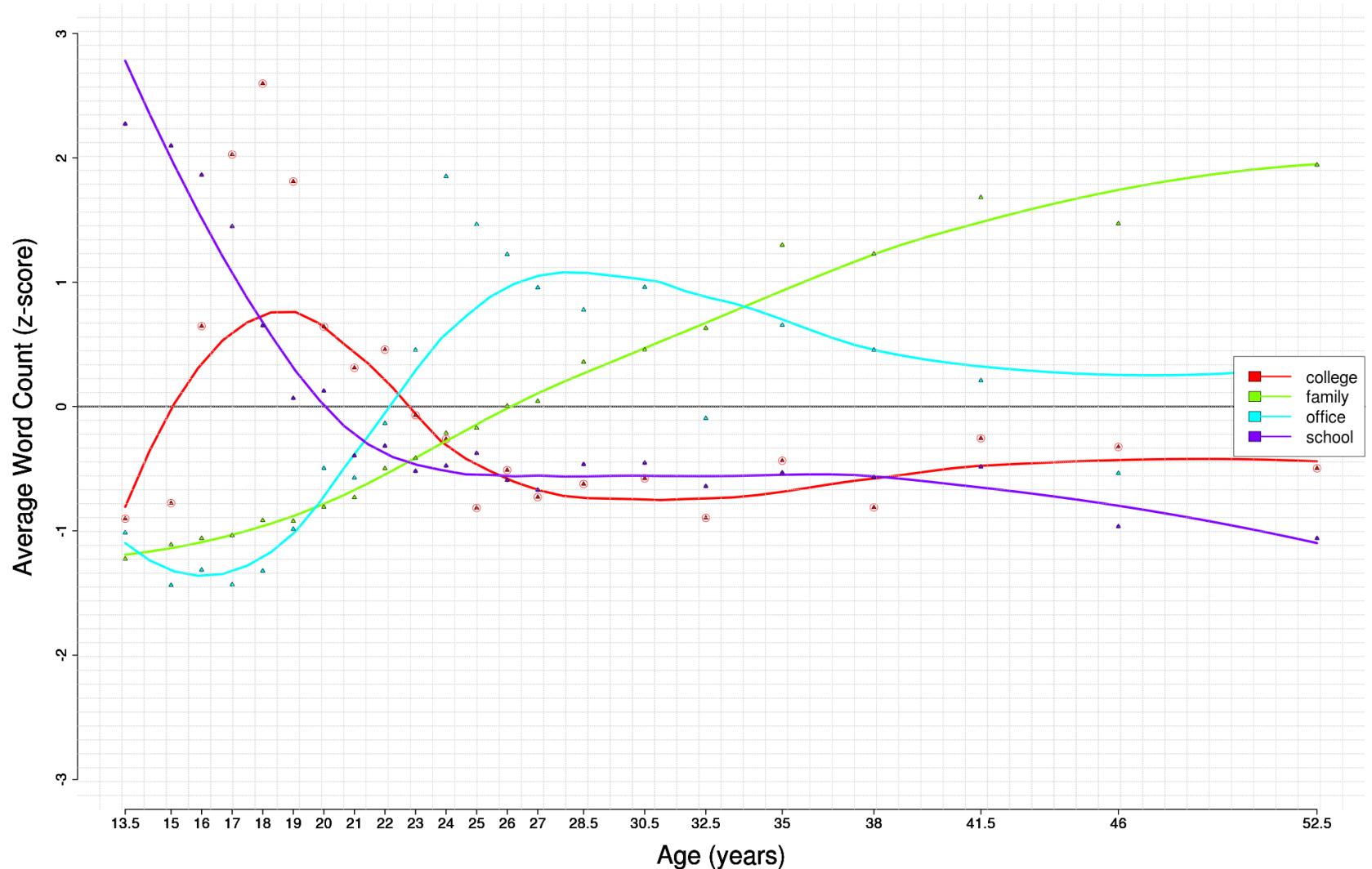
nature sense sound totally alive reality
where shall there's must completely
own evil didn't thoughts light
the_sky deep isn't dead listening
less blood humanity moon I'm_going:_ apparently I'd they're perhaps
like_a though you_know sea words realized
darkness become dream seems doesn't
exists right_now_ writing zombie those
upon quite epic what's dreams entire myself
through sigh death die poetry I'll soul
itself I_don't dark you're human kind
religion three philosophy society reading
awesome book it's

WORD USE ACROSS LIFE SPAN

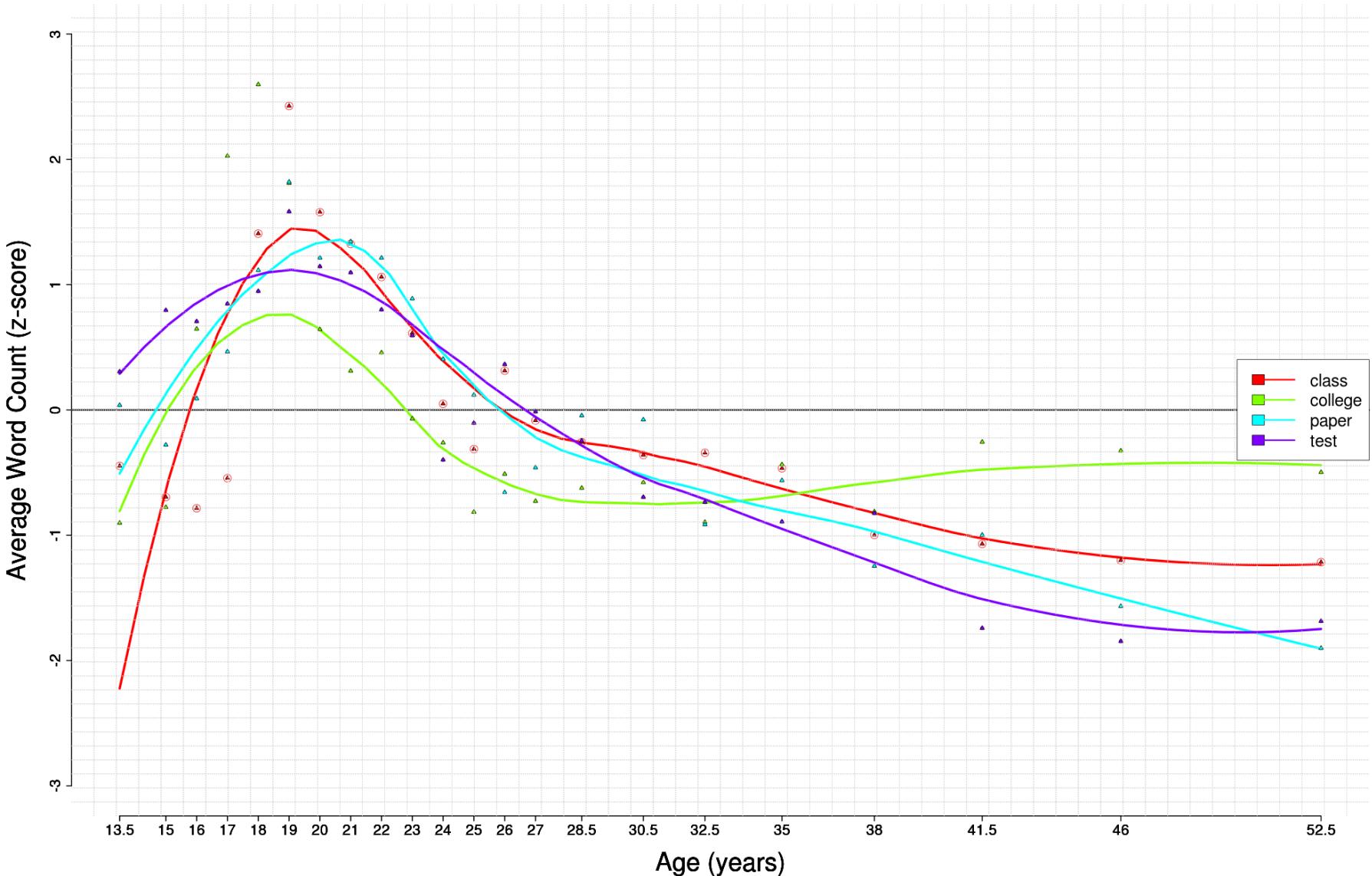
Activities and concerns .1

Orientation towards others .2

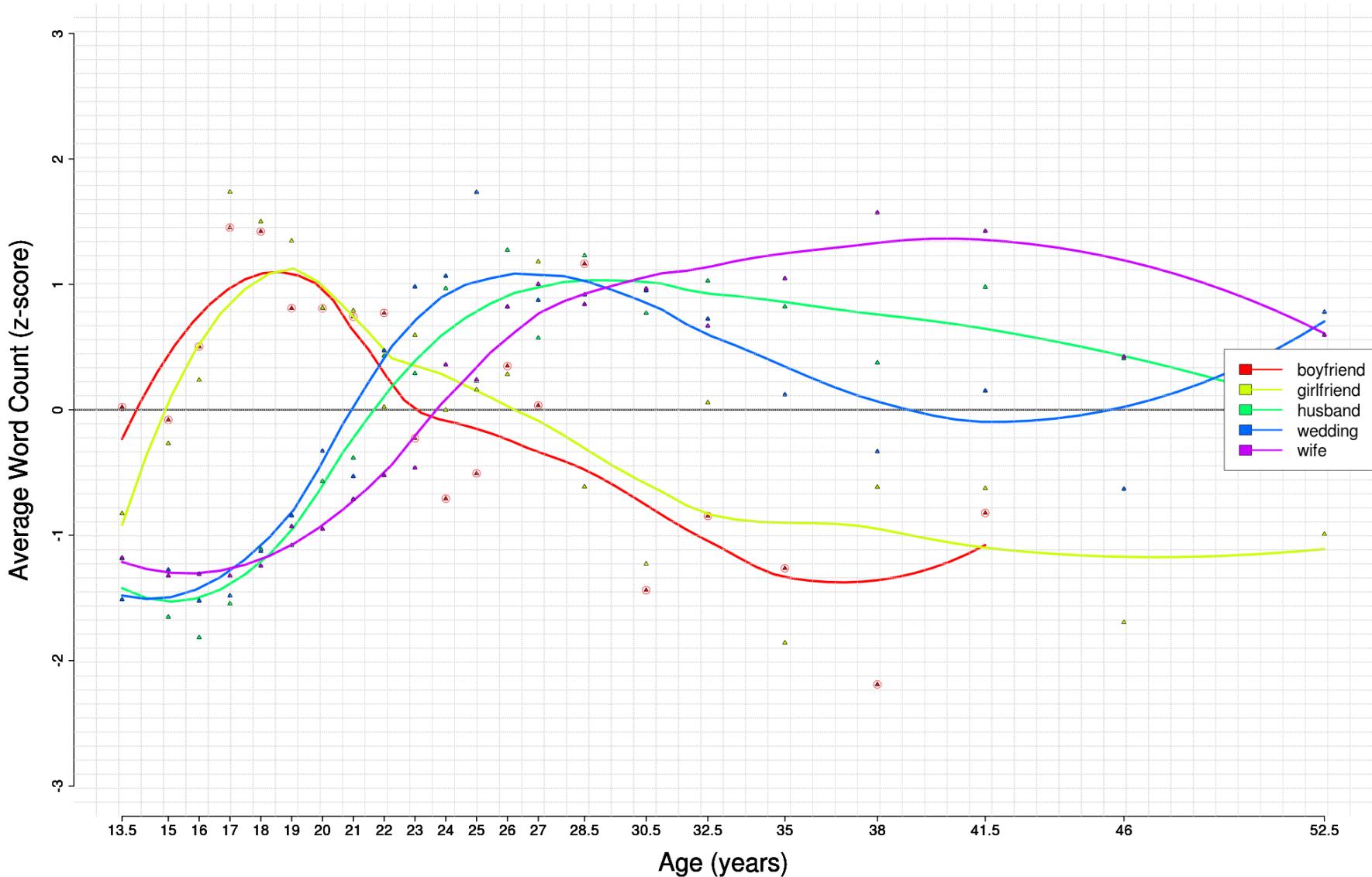
Emotional states and attitudes .3



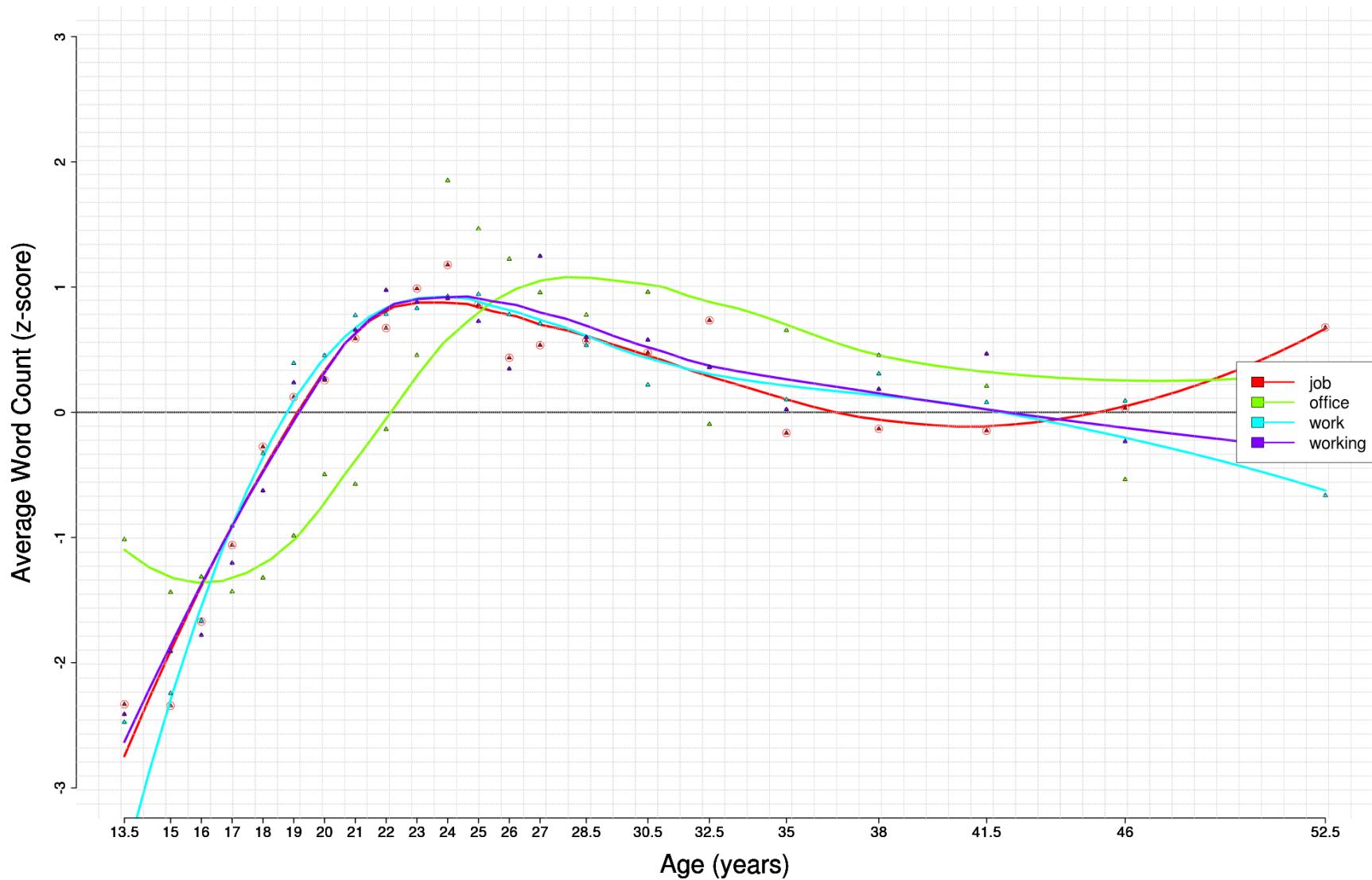
college



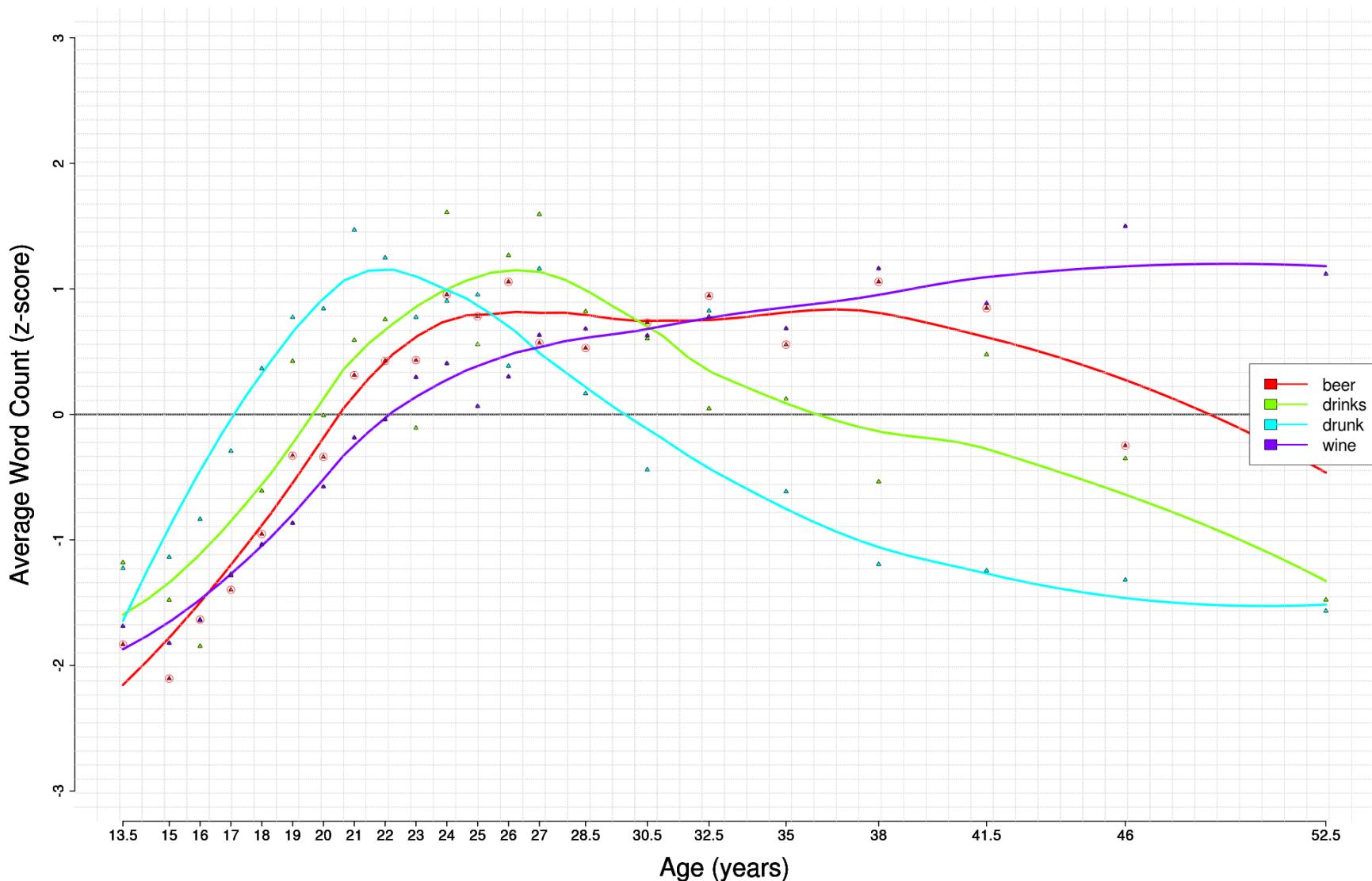
partnering



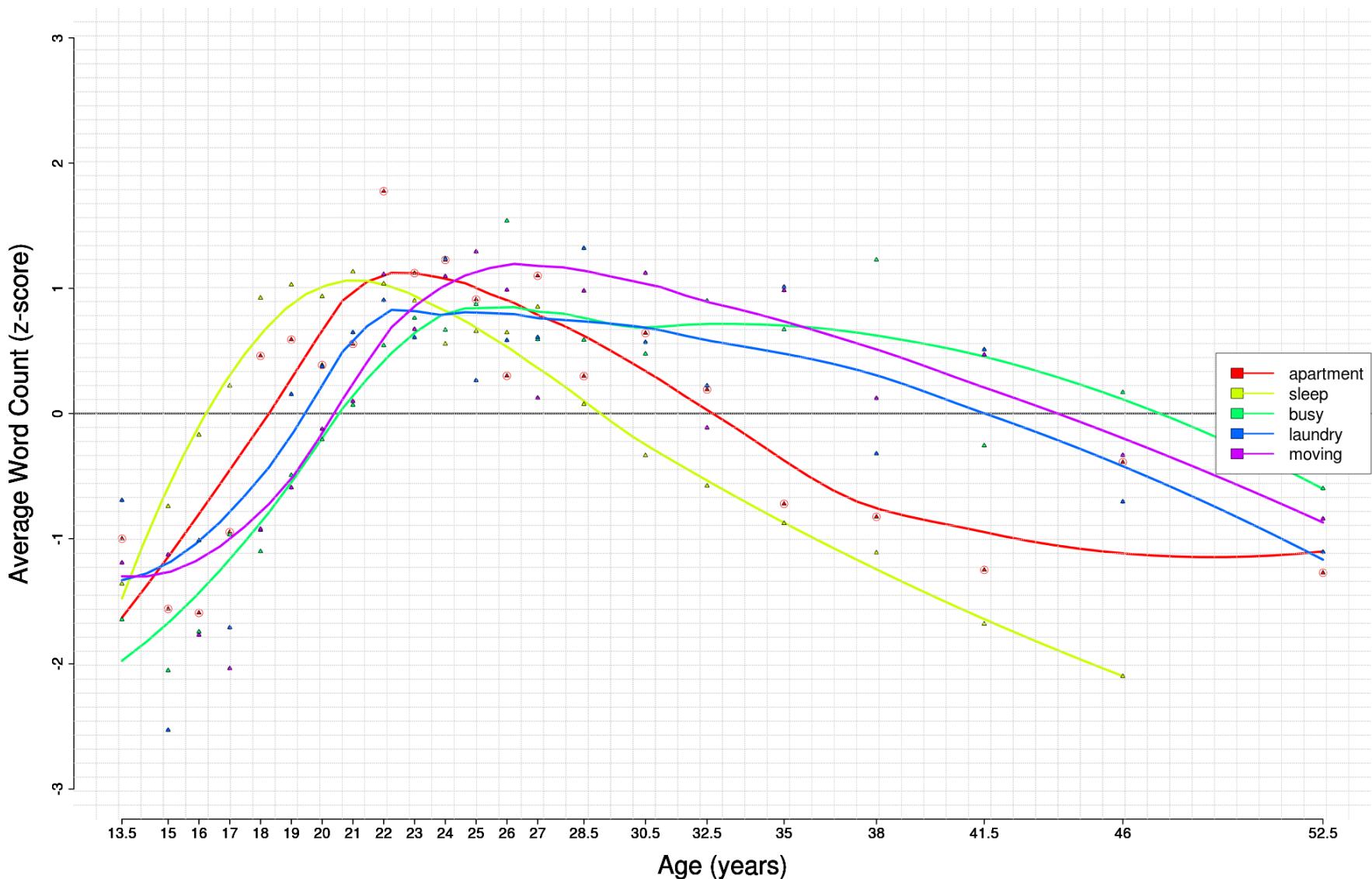
work



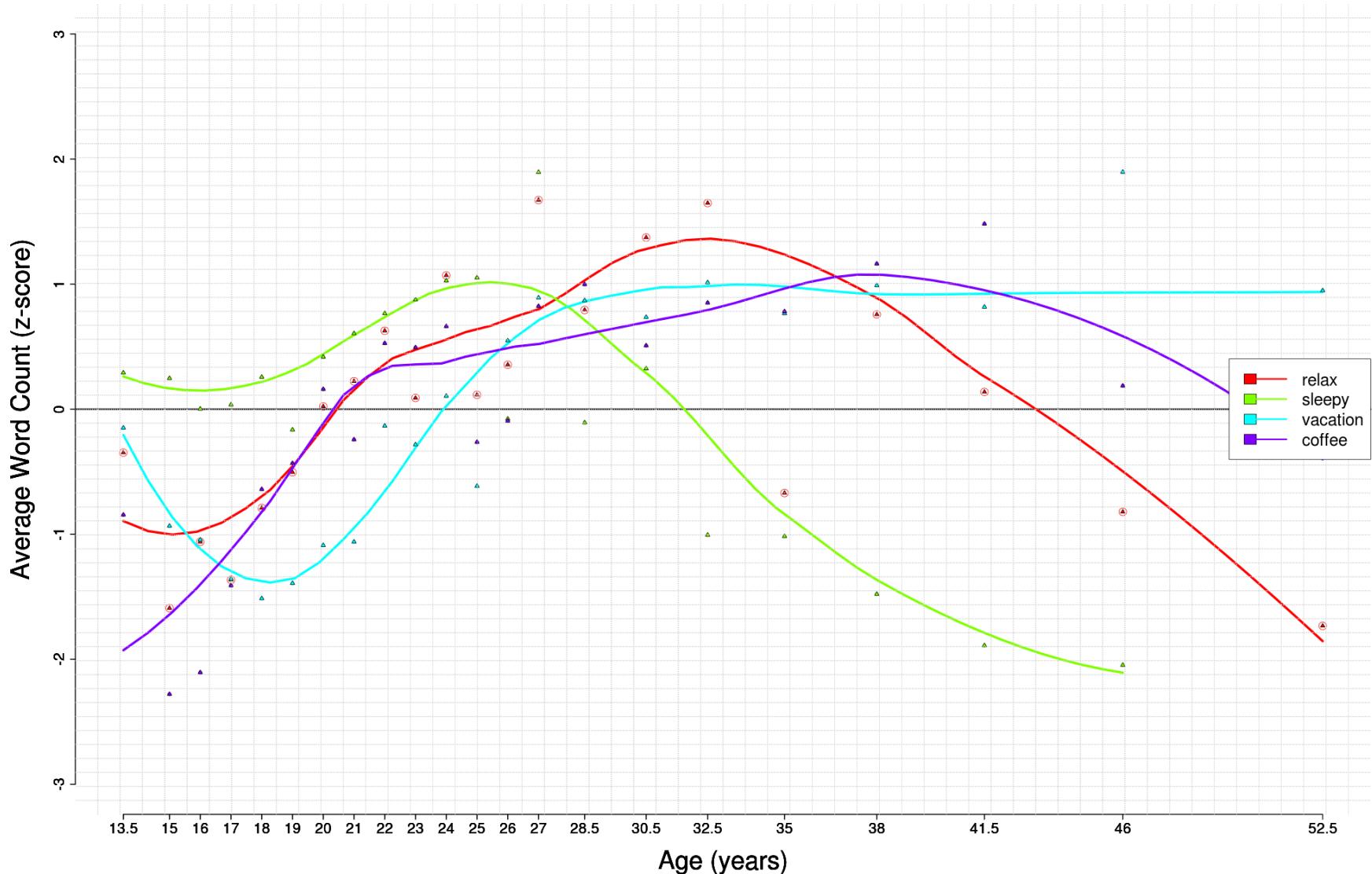
alcohol consumption



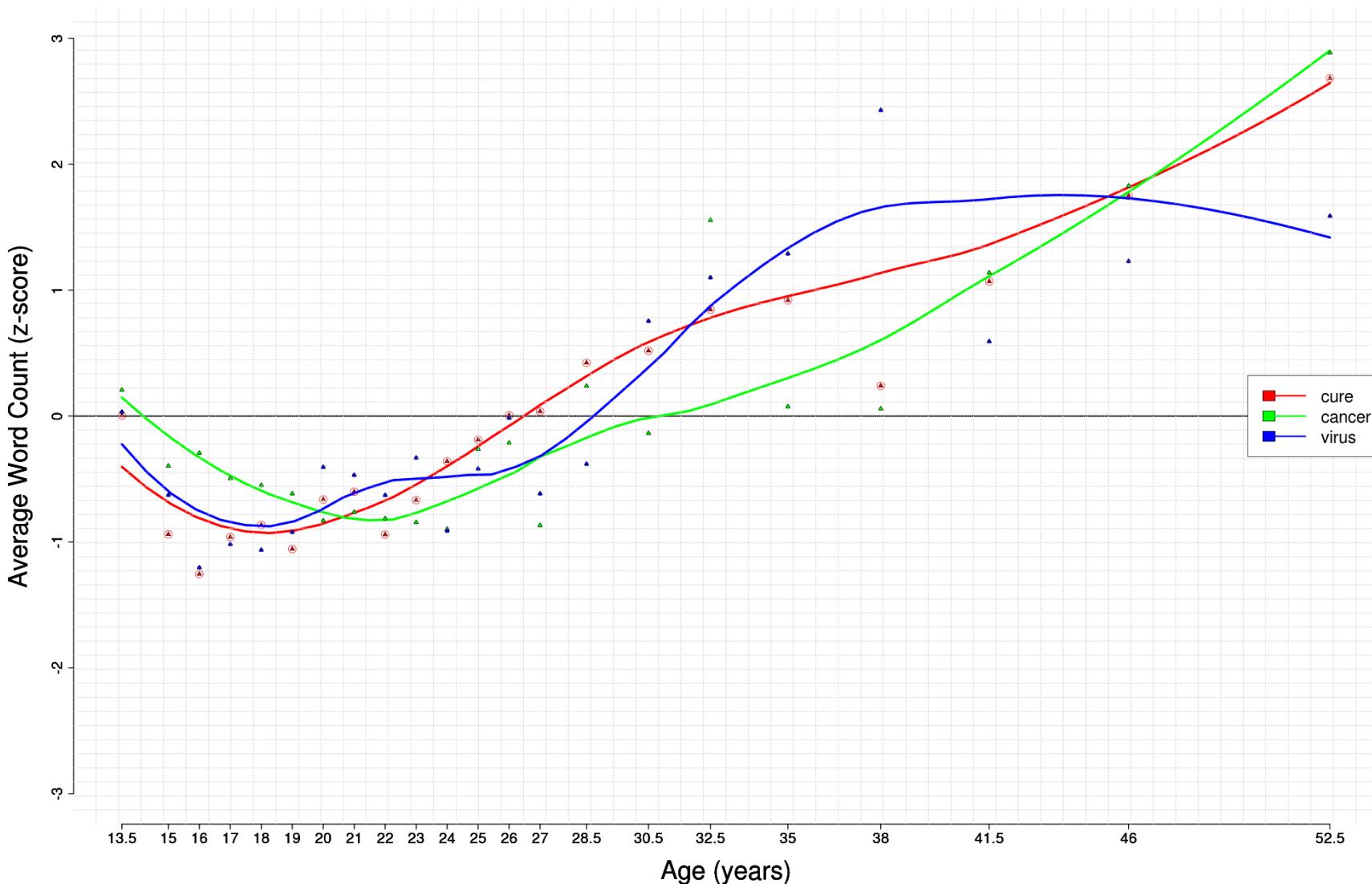
grown-up life

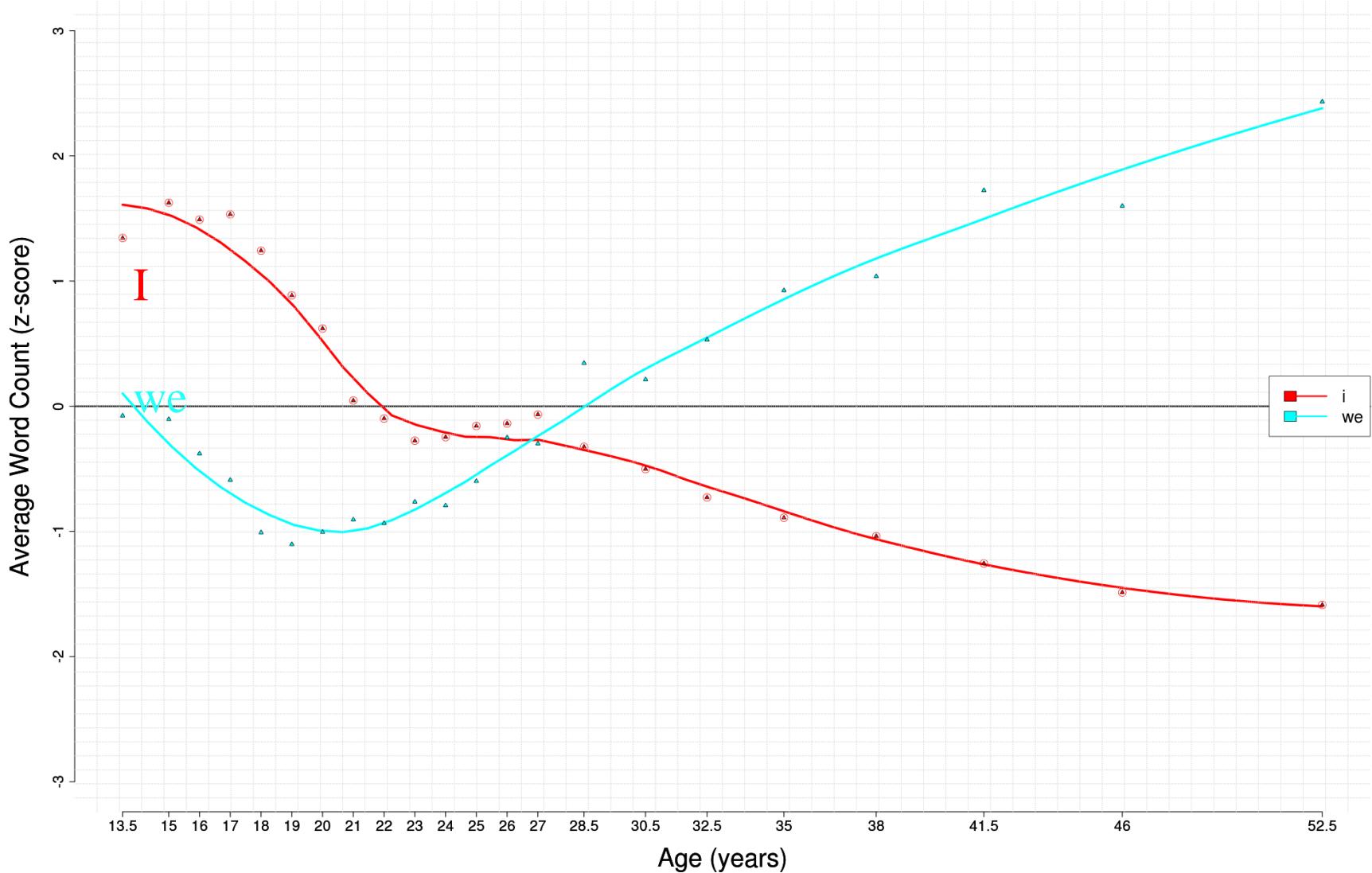


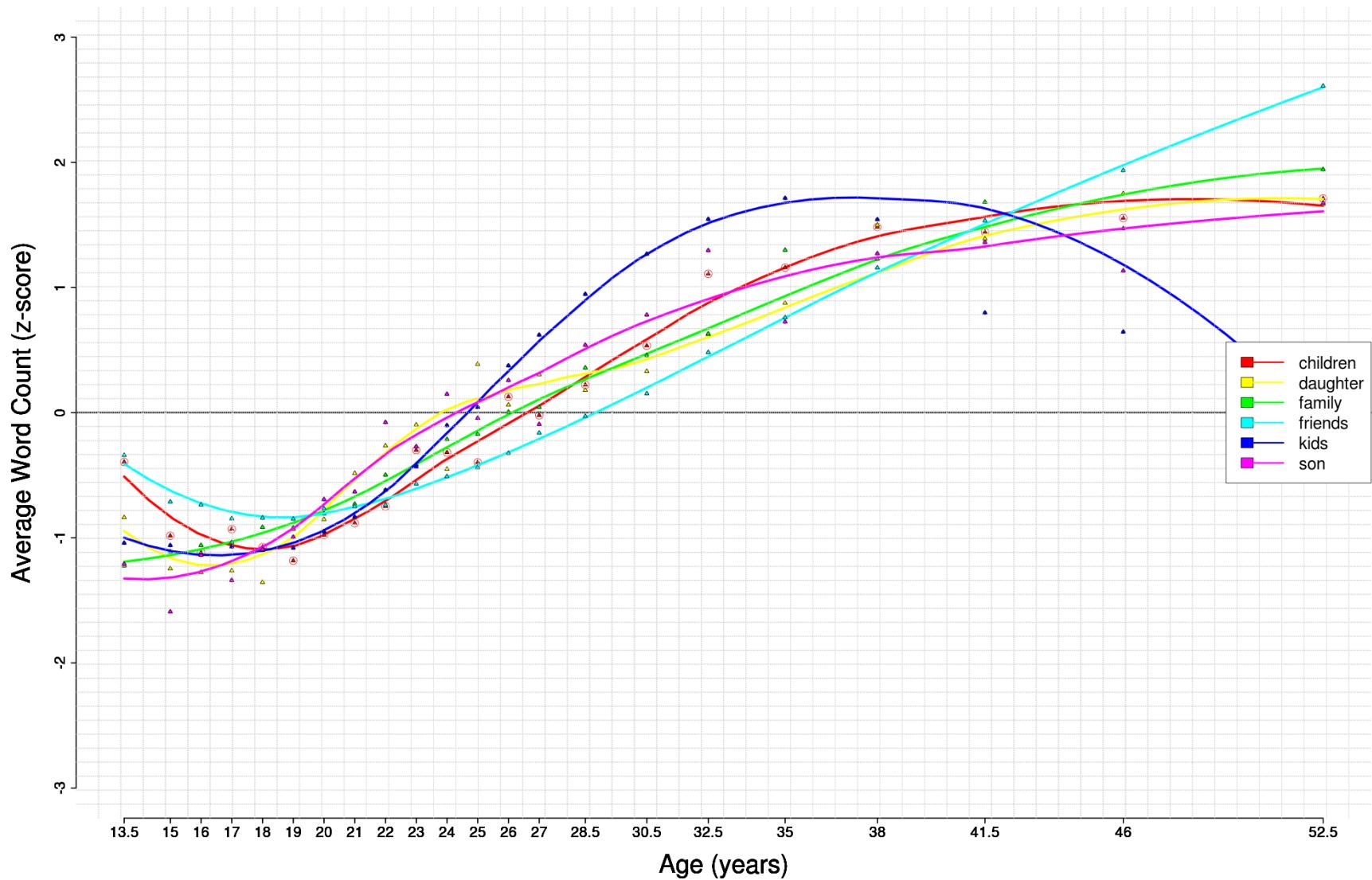
relaxation



diseases







Conclusions

- People's words on Facebook reveal their personality and concerns
 - Combining with online *surveys* adds huge power
- Word use correlates with Gallup polls and CDC data
 - Twitter: more data, better spatial and temporal resolution
- Open vocabulary allows hypothesis generation
 - Enhanced by ambiguity resolution, phrase detection, topic analysis, and visualization