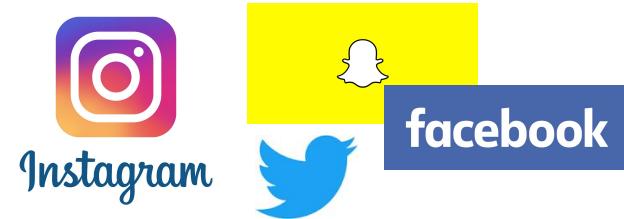


Personality Traits and Social Media Behavior

Final presentation

Tomasz Kolonko, Mario Kaufmann

Introduction



- Social media platforms very widely used
 - Facebook: 1.87 billion active users in 2017 [1]
- Users disclose personal information
 - Friends
 - Interests
 - Opinions
- Attempts are made to use that information for various purposes

Project Goal

- Investigate relationship between personality and
 - Social network properties
 - Disclosure behavior
- Find meaningful correlations
- Make machine learning based predictions
 - Predict personality from social network
 - Predict social network / behavior from personality

Dataset



- Dataset from myPersonality project [2]
- 10'000 Facebook status updates of 250 users
- Personality features after Big Five
- Social network properties

Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	#AUTHID	STATUS	sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	cCON	cOPN	DATE
2	b7b7764cfa1c523e4e93ab2a79a946c4	likes the sound of thunder.	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/19/09 03:21 PM
3	b7b7764cfa1c523e4e93ab2a79a946c4	is so sleepy it's not even funny t►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/02/09 08:41
4	b7b7764cfa1c523e4e93ab2a79a946c4	is sore and wants the knot of mu►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/15/09 01:15 PM
5	b7b7764cfa1c523e4e93ab2a79a946c4	likes how the day sounds in this ►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/22/09 04:48 AM
6	b7b7764cfa1c523e4e93ab2a79a946c4	is home. <3	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/20/09 02:31 AM
7	b7b7764cfa1c523e4e93ab2a79a946c4	www.thejokerblogs.com	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/16/09 03:21 PM
8	b7b7764cfa1c523e4e93ab2a79a946c4	saw a nun zombie, and liked it. ►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/27/09 05:41 AM
9	b7b7764cfa1c523e4e93ab2a79a946c4	is in Kentucky. 421 miles into he►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/18/09 06:34 AM
10	b7b7764cfa1c523e4e93ab2a79a946c4	was about to finish a digital pain►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/09/09 14:58
11	b7b7764cfa1c523e4e93ab2a79a946c4	is celebrating her new haircut by►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/07/09 23:41
12	b7b7764cfa1c523e4e93ab2a79a946c4	has a crush on the Green Lanter►	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/15/09 07:48 PM

N	O	P	Q	R	S	T
NETWORKSIZE	BETWEENNESS	NBETWEENNESS	DENSITY	BROKERAGE	NBROKERAGE	TRANSITIVITY
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1

Big Five Personality Traits

- Five broad dimensions to describe human personality [3,4]
 - Openness to experience: *inventive / curious* vs. *consistent / cautious*
 - Conscientiousness: *efficient / organized* vs. *easy-going / careless*
 - Extraversion: *outgoing / energetic* vs. *solitary / reserved*
 - Agreeableness: *friendly / compassionate* vs. *analytical / detached*
 - Neuroticism: *sensitive / nervous* vs. *secure / confident*

Social Network Properties

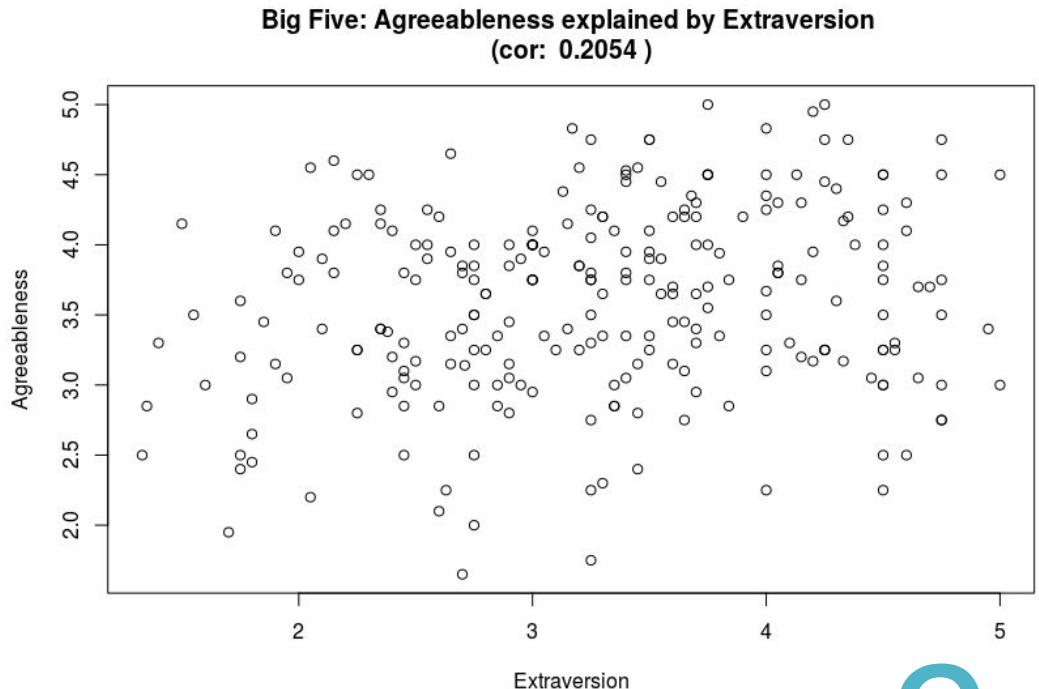
- Multiple properties of the user's social network:
 - Network size: number of friends
 - Betweenness: shortest path connections through user
 - Density: ratio of user's connections to connections within network

Correlations: Big Five

- No correlation was found between the different personality traits
 - The five personality traits should not correlate strongly since they describe distinct personality dimensions

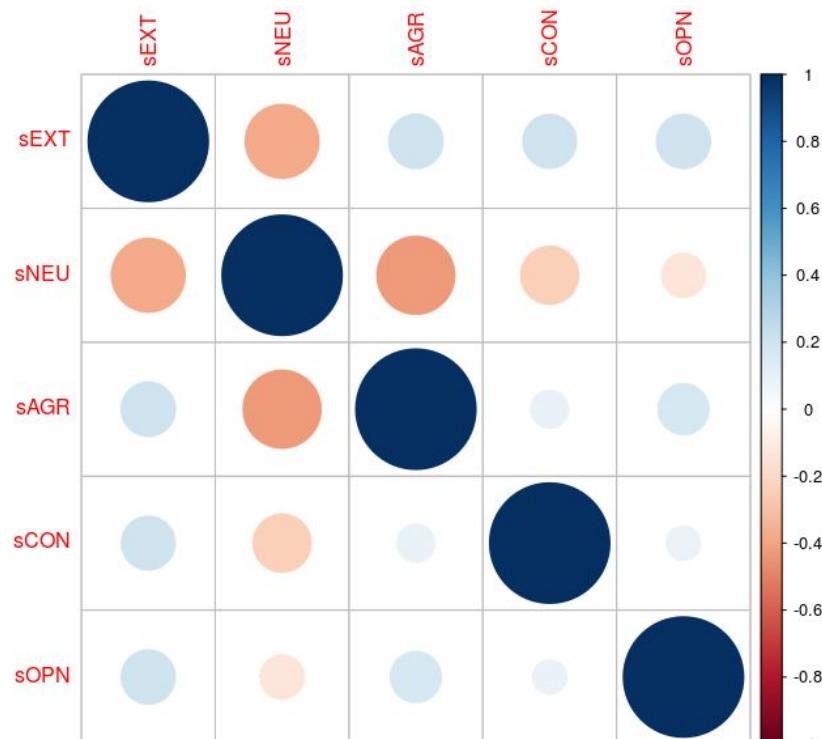
Correlations: Big Five

- Here Agreeableness is explained by Extraversion with a correlation of 0.2
- Correlations below 0.5 can be seen as insignificant



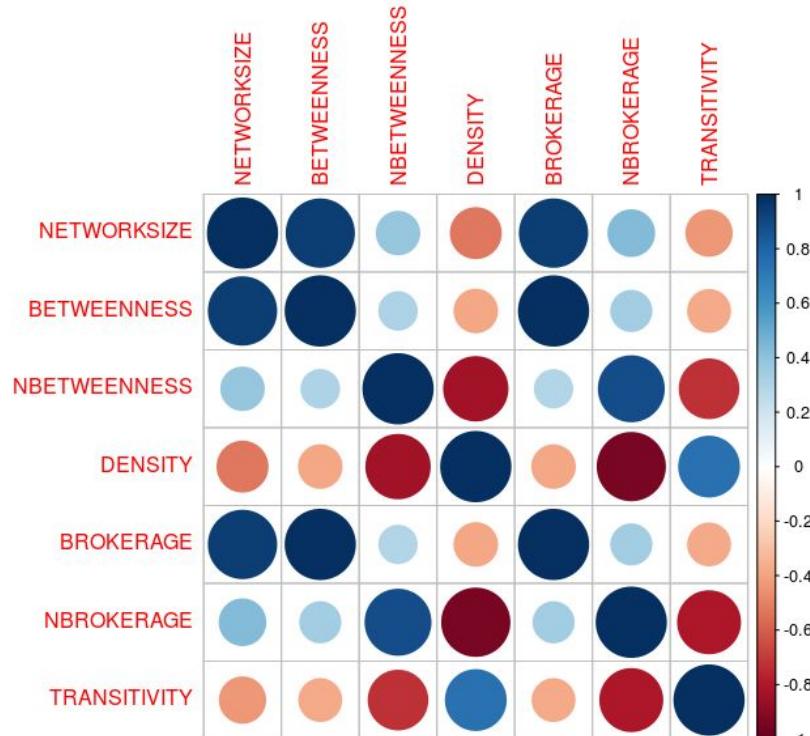
Correlations: Big Five

- This matrix shows the correlations between all the personality traits
- Neuroticism correlates weakly with Extraversion (-0.377)
- Neuroticism correlates weakly with Agreeableness (-0.42)



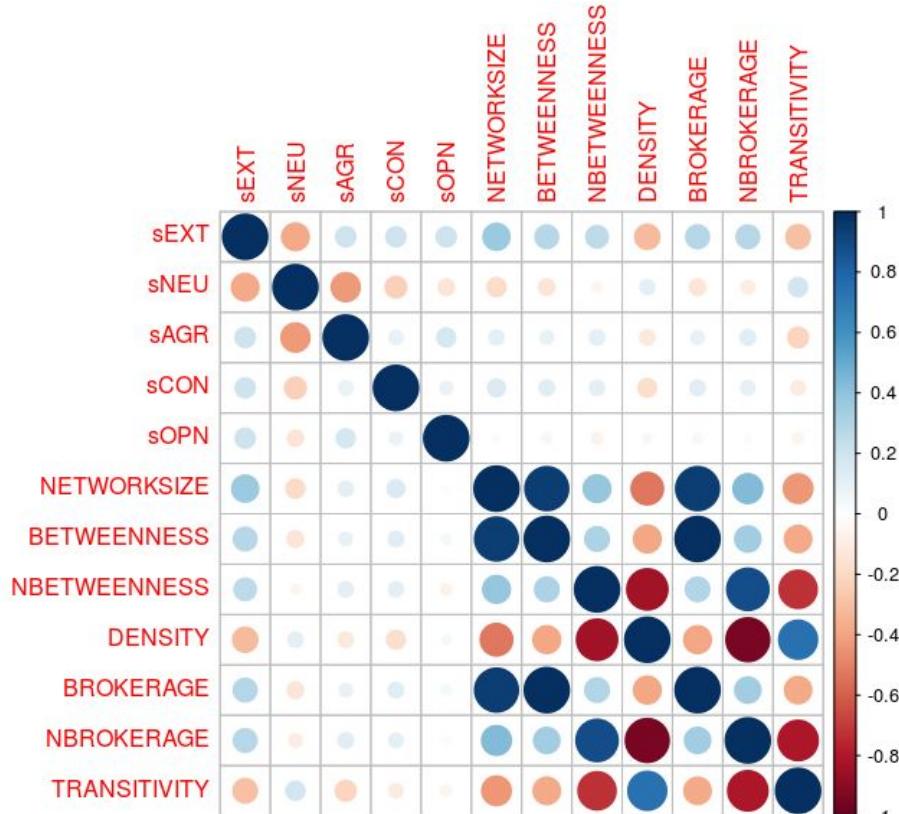
Correlations: Ego Networks

- Correlations between the Ego Network parameters are pretty strong
- But how do they correlate with the personality traits from Big Five?



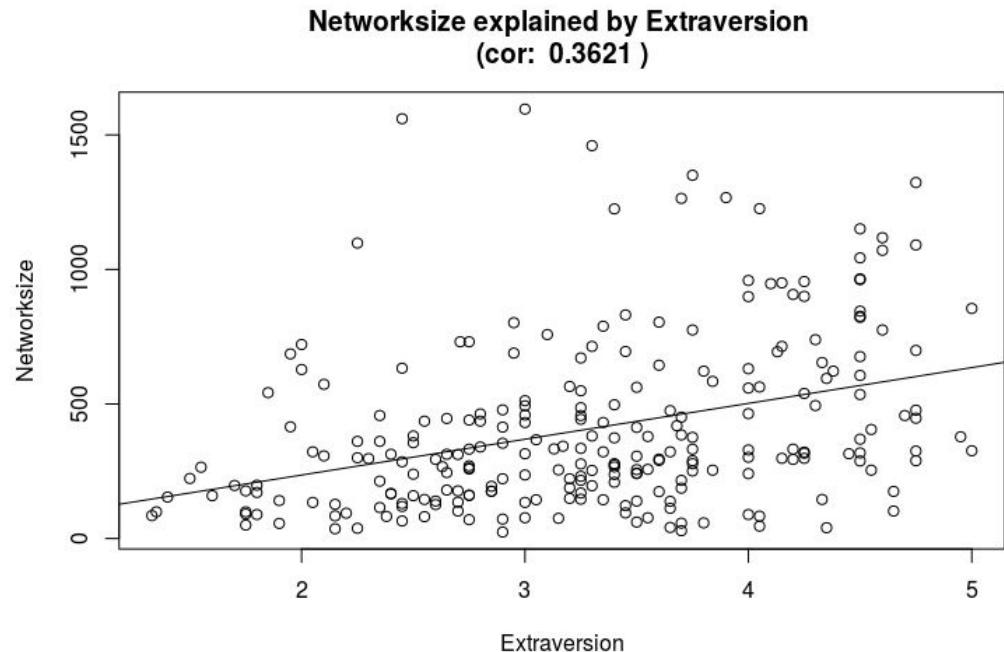
Correlations: Big Five - Ego Networks

- Not so well ;(



Correlations: Big Five - Ego Networks

- Need to find other parameters from our dataset

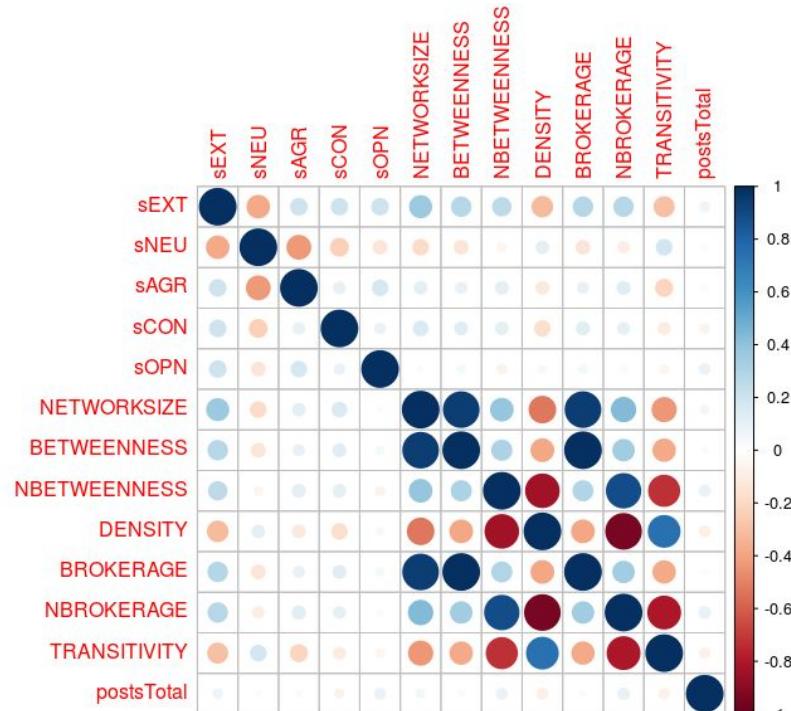


Creating new Variables

- Since no obvious correlations were found we try to get new information out of the existing by:
 - Aggregating different variables into new ones
 - Introducing new variables for the number of posts and when they were written (morning vs. evening)
 - Introducing new variables that describe the posts (length, exclamation marks, capital letters...)

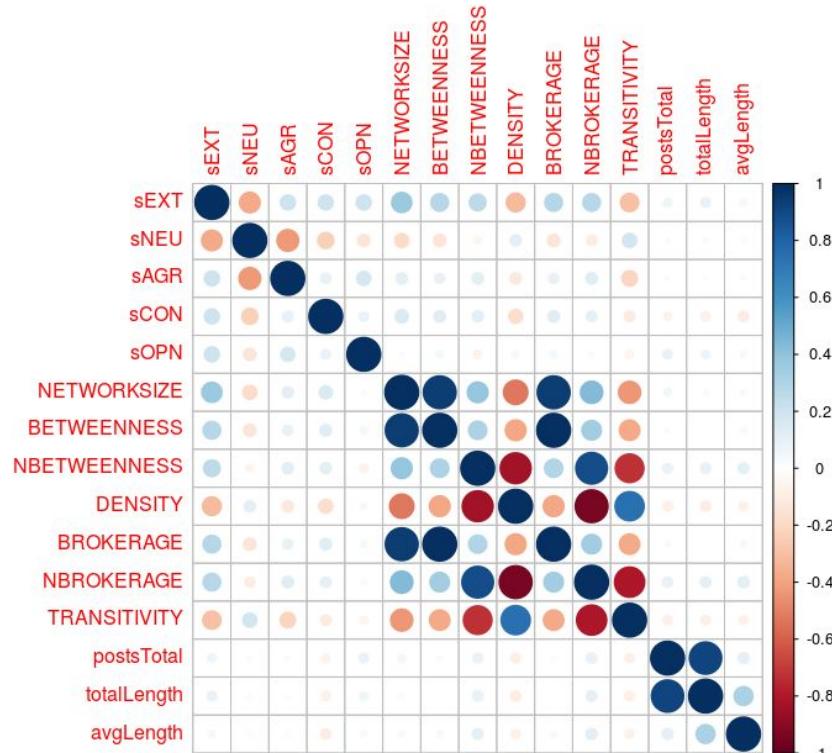
Number of Posts per User

- We hoped to find a correlation between the number of posts a user makes and her extraversion or openness...



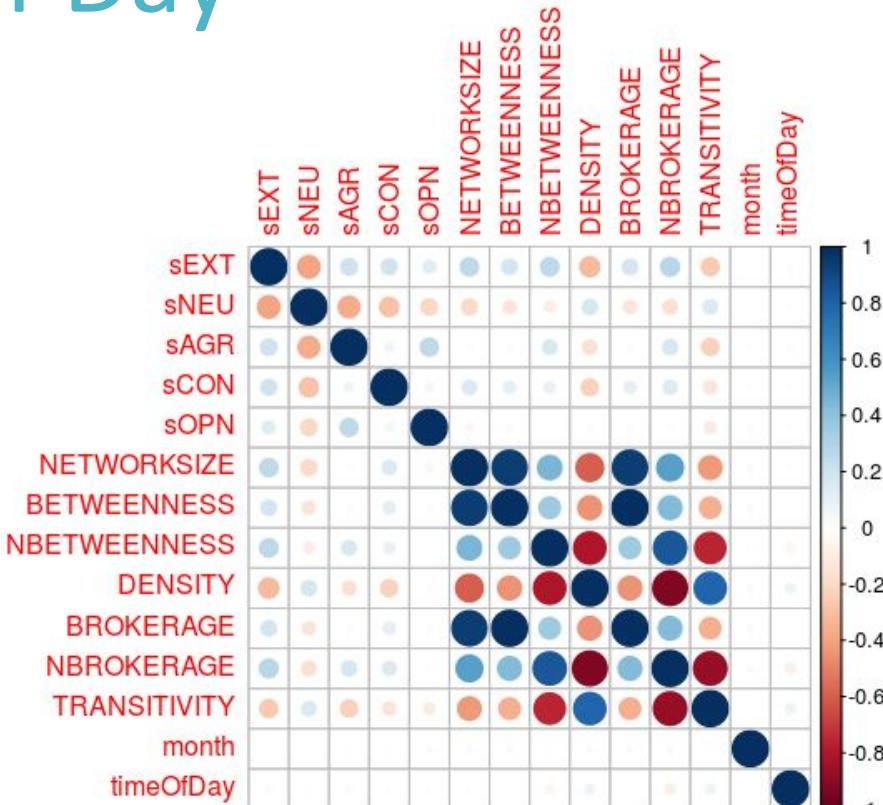
The Length of Messages

- We hoped to find a correlation between the total length of all posts or the average length per post of a user and his extraversion or openness...



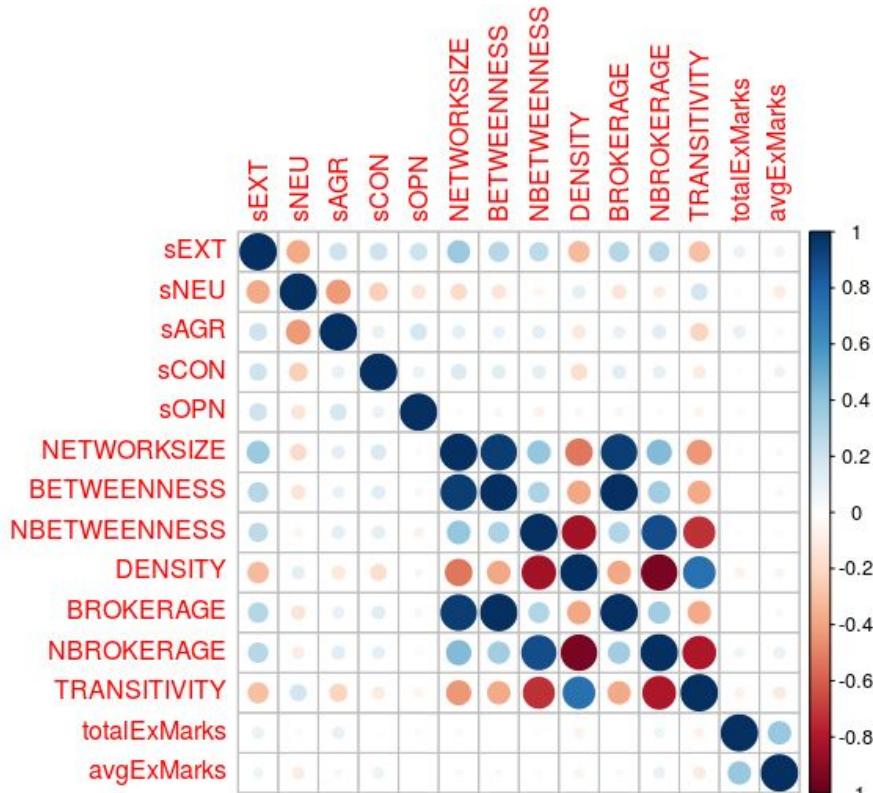
Month and Time of Day

- We hoped to find a correlation of a Big5 factor and the time of year, or time of day...



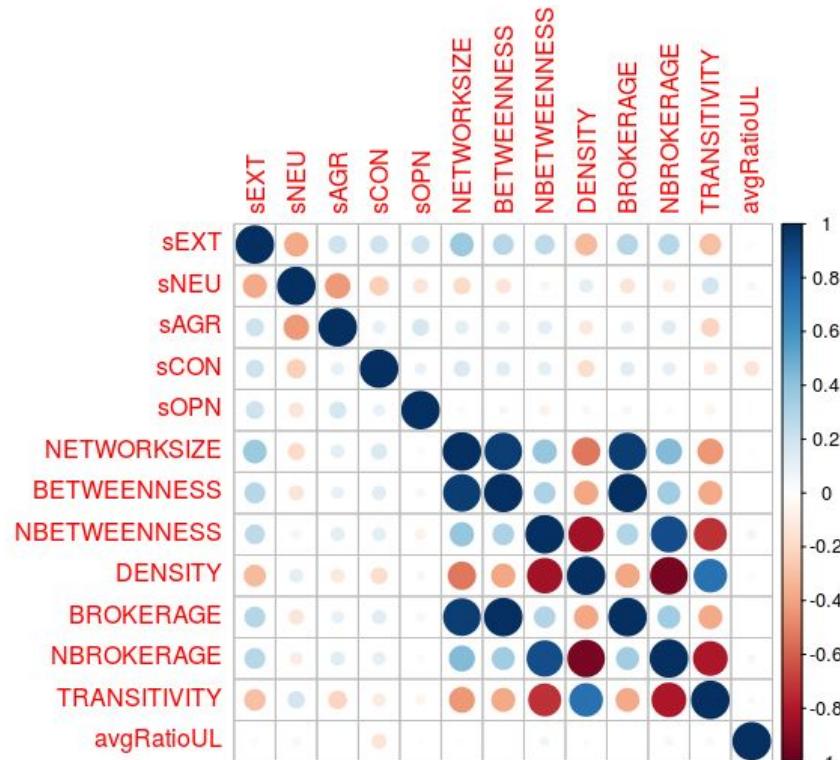
The Number of Exclamation Marks

- We hoped to find a correlation of some Big5 factors and the total or average of used exclamation marks...



Upper and Lower Case Letters

- We hoped to find a correlation of some Big5 factors and the ratio of upper and lower case letters



Multiple Linear Regression

With the Multiple Linear Regression model several predictors are used in order to predict the y-value:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{i1} + \hat{\beta}_2 \cdot x_{i2} + \dots + \hat{\beta}_p \cdot x_{ip} + \epsilon_i$$

Multiple Linear Regression

- 28.43% of Neuroticism can be explained by the predictors Agreeableness, Extraversion, Conscientiousness, NBetweenness and Transitivity
- 28.37% of Extraversion can be explained by Neuroticism, Networksize, Brokerage, Openness and NBetweenness

Machine Learning

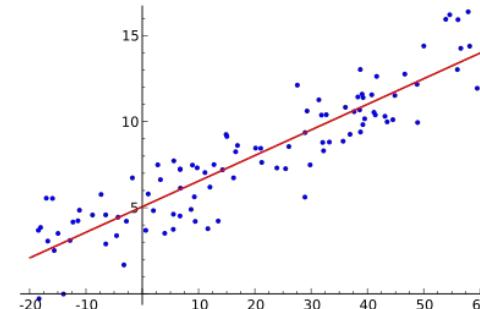
- Use of machine learning rapidly growing
- Is machine learning applicable here?

Possibility:

predict user's personality and from that preferences
(recommender systems or advertisement)

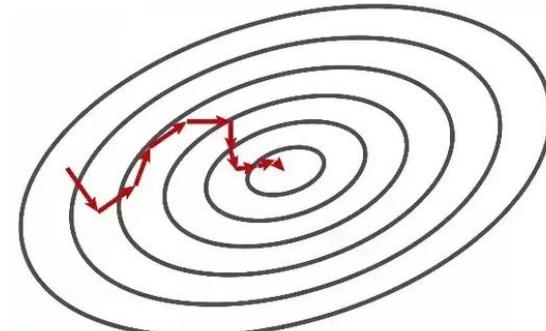
Machine Learning: Linear Regression

- Predict output value with continuous domain
 - Personality features on a scale of 1 to 5
 - Network properties
- Model linear relationship by training regression model
 - Predict values
 - Compute error
 - Minimize error



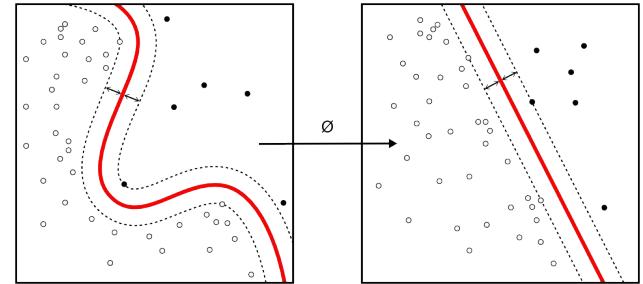
Machine Learning: Linear Regression

- Minimize the error by applying stochastic gradient descent
- Take step towards optimum for every sample:
 - Consider sample label
 - Adjust parameters to optimize for sample
 - Save new parameters



Machine Learning: SVM

- Predict discrete output values
 - Binary personality features
- Train classification model
 - Assign label to every training sample
 - Achieve a clear separation of the two classes
- With kernel trick nonlinear relationships can be modelled



Experiment Setup

- Predict personality features from network properties
 - Input data: network size, nBetweenness, density, nBrokerage, transitivity
 - Output labels: extraversion and agreeableness
- Also: predict network properties from personality

Experiment Setup

- Split up data set into training and test set (70% / 30%)
- Train on training set, validate performance on test set
- Remaining parameters manually tuned and set by hand
 - SVM: regularization parameter, kernel parameters
 - Linear regression: number of iterations, learning rate

Results: Network -> Personality

- Extraversion:
 - SVM: median accuracy of 65.83%
 - Linear regression: median cost of 0.677 per sample
- Agreeableness:
 - SVM: median accuracy of 56.1%
 - Linear regression: median cost of 0.476 per sample

Results: Personality -> Network

- SVM could not be used (only continuous values available)
- Overall huge errors
- No useable prediction possible
- Reasons:
 - high variance in data
 - relatively little data

Dataset - Shortcomings

- There are only 250 users with 9'917 posts
- Posts are unevenly distributed among the 250 users
- Only 19 variables of which 7 are highly dependent
- Very high redundancy
- If redundancy is reduced only 250 data points remain for machine learning algorithms which is way to few

Dataset - Shortcomings

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	#AUTHID	STATUS	sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	cCON	cOPN	DATE
2	b7b7764cf1c523e4e93ab2a79a946c4	likes the sound of thunder.	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/19/09 03:21 PM
3	b7b7764cf1c523e4e93ab2a79a946c4	is so sleepy it's not even funny t	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/02/09 08:41
4	b7b7764cf1c523e4e93ab2a79a946c4	is sore and wants the knot of mu	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/15/09 01:15 PM
5	b7b7764cf1c523e4e93ab2a79a946c4	likes how the day sounds in thi	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/22/09 04:48 AM
6	b7b7764cf1c523e4e93ab2a79a946c4	is home. <3	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/20/09 02:31 AM
7	b7b7764cf1c523e4e93ab2a79a946c4	www.thejokerblogs.com	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/16/09 03:21 PM
8	b7b7764cf1c523e4e93ab2a79a946c4	saw a nun zombie, and liked it.	2.65	3	3.15	3.25	4.4	n	y	n	n	y	06/27/09 05:41 AM
9	b7b7764cf1c523e4e93ab2a79a946c4	is in Kentucky. 421 miles into he	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/18/09 06:34 AM
10	b7b7764cf1c523e4e93ab2a79a946c4	was about to finish a digital pain	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/09/09 14:58
11	b7b7764cf1c523e4e93ab2a79a946c4	is celebrating her new haircut by	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/07/09 23:41
12	b7b7764cf1c523e4e93ab2a79a946c4	has a crush on the Green Lanter	2.65	3	3.15	3.25	4.4	n	y	n	n	y	07/15/09 07:48 PM

N	O	P	Q	R	S	T
NETWORKSIZE	BETWEENNESS	NBETWEENNESS	DENSITY	BROKERAGE	NBROKERAGE	TRANSITIVITY
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1
180	14861.6	93.29	0.03	15661	0.49	0.1

Dataset - Shortcomings

9881	d38a81dfbfbbd5f2a2a0a03e9db304c6	Kaunist s_brap_eva k_igile! :)	1.85	3.55	3.45	3.4	3.45	n	y	n	n	n	02/14/10 11:34 AM
9882	ddaed24e83f0f9958336b52cf7a89373	In Canada with fam	4.25	2.5	3.25	4.25	4.5	y	n	n	y	y	12/31/09 10:42 PM
9883	ddaed24e83f0f9958336b52cf7a89373	Thank you all. God bless!	4.25	2.5	3.25	4.25	4.5	y	n	n	y	y	02/12/10 03:40
9884	6980ce18350d98916f56c95b4dc4496d	Merry X-Mas!	2.75	2.8	3.25	2.5	4.35	n	y	n	n	y	12/25/09 06:18 PM
9885	6980ce18350d98916f56c95b4dc4496d	is moving today. I've had to mov	2.75	2.8	3.25	2.5	4.35	n	y	n	n	y	01/30/10 07:27 PM
9886	c255a1cb2939ce6b4719a8a0cc085624	black	4.75	3	2.75	4.5	4.75	y	y	n	y	y	01/08/10 01:50
9887	8974aab97d9fc4e3a53ba126b5eedd81	Thanks for all the Birthday and O	3.2	2.9	3.25	2.7	4.05	n	y	n	n	y	01/09/10 01:01
9888	35efb99775d5ee7e83cf7912591984d5	Facebook me marea. Me hates i	2.45	4	2.85	2.35	4.1	n	y	n	n	y	01/18/10 09:15 AM
9889	a764aca41dca158d7a191505dcc8ce47f	Red	3.7	2.5	4.2	4.1	3.6	y	n	y	y	n	01/11/10 04:19
9890	deb899e426c1a5c66c24eeb0d7df6257	About mornings and winter, and n	2.15	2.15	4.1	2.9	4.6	n	n	y	n	y	01/21/10 10:04 PM
9891	deb899e426c1a5c66c24eeb0d7df6257	little things give you away.	2.15	2.15	4.1	2.9	4.6	n	n	y	n	y	01/28/10 11:48 AM
9892	ea28a927cb6663480ea33ca917c3c8ba	is wishing it was Saturday.	4.05	3.35	3.8	3.95	4.5	y	y	y	y	y	02/02/10 20:08
9893	ea28a927cb6663480ea33ca917c3c8ba	is studying hard for the G.R.E.	4.05	3.35	3.8	3.95	4.5	y	y	y	y	y	02/09/10 20:04
9894	5532642937eb3497a43e15dbb23a9d2d	snipers get more head	1.4	4.05	3.3	3.4	3.95	n	y	n	n	y	02/10/10 02:01
9895	a286b7286b1247d4a7851709e9f31e1e	Last night was amazing! Not only	4.25	3	3.25	3.5	4	y	y	n	y	y	02/08/10 18:28

Dataset - Improvements

In order to have more meaningful and reliable predictions we propose the following improvements for future datasets:

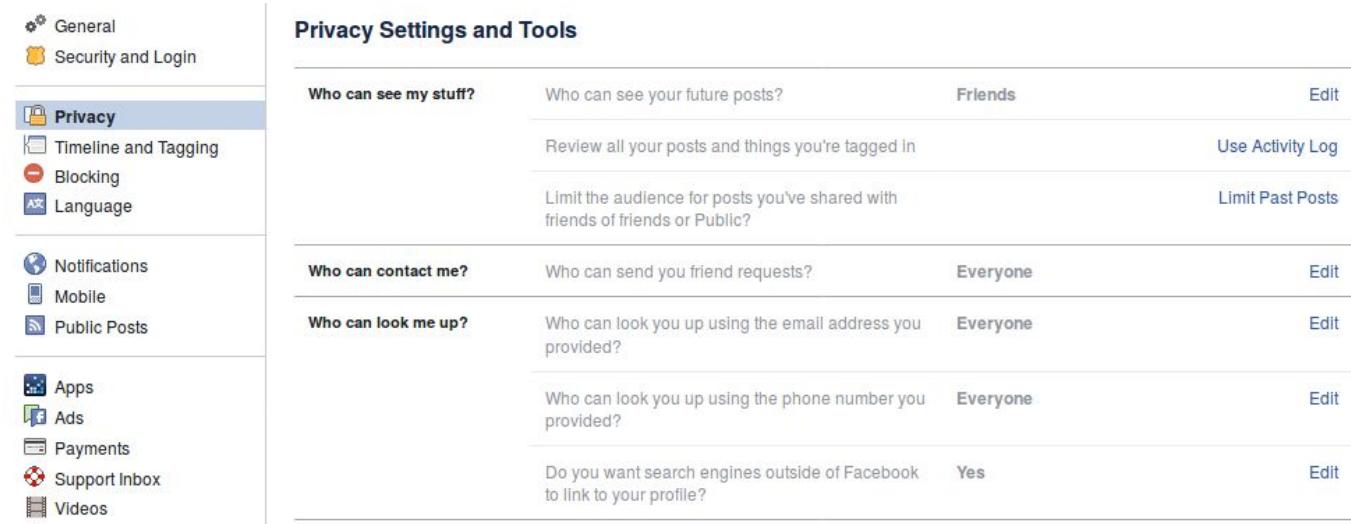
- Much larger data set
- If posts and date fields are important remove users with fewer than 48 posts per year (7022 posts by 75 users)
- More relevant data needed

Dataset - Relevant Data to Big5

- Studies show gender plays a big role [5,6,7]
- Technical Information like logins, all privacy settings and general settings

Dataset - Relevant Data to Big5

- Studies show gender plays a big role [5,6,7]
- Technical Information like logins, all privacy settings and general settings



The screenshot shows the 'Privacy Settings and Tools' section of a Facebook profile. On the left, there's a sidebar with various privacy-related tabs: General, Security and Login, Privacy (which is selected and highlighted in blue), Timeline and Tagging, Blocking, Language, Notifications, Mobile, Public Posts, Apps, Ads, Payments, Support Inbox, and Videos.

Privacy Settings and Tools

Setting	Current Value	Action
Who can see my stuff?	Friends	Edit
Who can contact me?	Everyone	Edit
Who can look me up?	Everyone	Edit
Do you want search engines outside of Facebook to link to your profile?	Yes	Edit

For the 'Who can see my stuff?' setting, the current value is 'Friends'. There are two options listed: 'Review all your posts and things you're tagged in' and 'Limit the audience for posts you've shared with friends of friends or Public?'. Both have 'Edit' links next to them.

For the 'Who can contact me?' setting, the current value is 'Everyone'. There are two options listed: 'Who can send you friend requests?' and 'Who can look you up using the email address you provided?'. Both have 'Edit' links next to them.

For the 'Who can look me up?' setting, the current value is 'Everyone'. There are two options listed: 'Who can look you up using the phone number you provided?' and 'Do you want search engines outside of Facebook to link to your profile?'. Both have 'Edit' links next to them.

At the bottom right of the table, there's a large 'Edit' button.

Dataset - Relevant Data to Big5

- Studies show gender plays a big role [5,6,7]
- Technical Information like logins, all privacy settings and general settings
- Geo-Location since studies show significant higher rates of psychological disorders in rural areas [8]
- Likes, Haha, Wow... and what did they like?

Dataset - Relevant Data to Big5

- Studies show gender plays a big role [5,6,7]
- Technical Information like logins, all privacy settings and general settings
- Geo-Location since studies show significant higher rates of psychological disorders in rural areas [8]
- Likes, Haha, Wow... and what did they like?



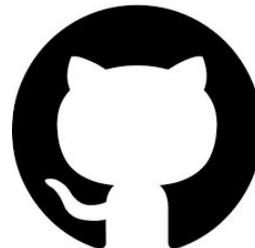
Dataset - Relevant Data to Big5

- Studies show gender plays a big role [5,6,7]
- Technical Information like logins, all privacy settings and general settings
- Geo-Location since studies show significant higher rates of psychological disorders in rural areas [8]
- Likes, Haha, Wow... and what did they like?
- Music, Movies, Series and attending Groups

Conclusion

- Dataset analysis:
 - No meaningful / very weak correlations found
 - Many ideas exist for creating new combinations and features
- Machine learning:
 - Prediction might be possible
 - Much more data with useable correlations needed

GitHub



<https://github.com/tomaszkolonko/WebAnalysisSeminar>

Questions

?

References

- [1] *Global social media research summary 2017*, Available:
<http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>
- [2] *myPersonality Project*, March 2017, Available: <http://mypersonality.org>
- [3] Goldberg, L. R. (1981). *Language and individual differences: The search for universals in personality lexicons*
- [4] https://en.wikipedia.org/wiki/Big_Five_personality_traits
- [5] Chapman, Benjamin P. et al., (2007). *Gender Differences in Five Factor Model Personality Traits in an Elderly Cohort: Extension of Robust and Surprising Findings to an Older Generation*
- [6] Weisberg, Yanna J., Colin G. DeYoung, and Jacob B. Hirsh. (2011). *Gender Differences in Personality across the Ten Aspects of the Big Five*
- [7] Costa Jr., Paul; Terracciano, Antonio; McCrae, Robert R. (2001). *Gender differences in personality traits across cultures: Robust and surprising findings*
- [8] N. T. Nguyen et al., (2005). *Personality Predicts Academic Performance: Exploring the moderating role of gender*