White Paper

Social Behavior & Privacy Disclosure on Facebook and Twitter

Akshata S. Salehittal akshata@umd.edu

Meghna Sardana Sarin msarin@umd.edu

Rahul Sarin sarinr@umd.edu

Introduction

With the prevalence and acceptance of Social Networking sites as one of the primary means for establishing and maintaining relationships, intentionally or inadvertently we share parts of our identity with the outside world. Millions of people routinely self-disclose personal information on social networking sites. While the level of self-disclosure may vary on several levels, gender may be commonly an observed criteria in the determining amount of self-disclosure. From expressing extremely personal feelings and opinions to documenting mundane details of everyday life, this type of public self-disclosure with multiple, wide and often ill-defined audiences blurs lines between publicness and privacy. It may also raise doubts and questions about self-disclosure in social media, and what are the factors that lead people to or prevent them from disclosing information.

When people join social networking sites, they begin by creating a profile. Several people may wish to create multiple profiles depending on the type of information they wish to share. They then make connections to existing friends, family, co-workers as well as those they meet through the site. A profile is a list of identifying information that could include your real name, some form of your real name or a pseudonym. Additionally, it also includes personal information like birthday, current location, hometown, ethnicity and personal interest. Members of such a platform, connect to others by sending a "friend request" if connecting on Facebook or "follow" a person if connecting on Twitter. In the first case, the request must be accepted by the other party to establish a connection. Friends can then access each other's profile though one may choose to limit what information is visible to the other by enforcing certain privacy settings. It is interesting to notice how different people behave on social networking sites. The platform gives one an opportunity to project an image of himself that is different and more appealing.

If we look into the relationship between gender and how comfortable a person is in revealing a certain type of information, there are a number of factors depending on what he may want to share. For example, women are known to have disclosed more about topics related to relationship, family, feelings, accomplishments etc. While, men on the other hand are known to have disclosed more about topics like cars, sports, and politics. However, disclosing other private and personal information may or may not be always related to gender. There are a number of other factors to consider. A person may not reveal his true identity in the fear of being misjudged by his potential or future employers, or may alter his posting habits to please the general public so as to be considered socially acceptable. Such factors should be taken into consideration when we wish to study the online posting behavior of people on social media.

This paper intends to discuss the online identity and disclosure issues among Facebook and Twitter users by means of analyzing the 2011 Facebook and Twitter Network Data dataset. It is interesting to analyze the online behavior of social media users towards disclosing confidential and private information, including their identity, to 'social media friends'. Through this analysis, we wish to obtain a general overview of users' outlook towards their online identity and the disclosure of information related to them. Some intriguing questions to ponder are the use of real names or pseudo names as usernames for user's Twitter profiles and whether there are gender gaps in such a behavior. Additionally, it was also of interest to us to find out what are some the factors associated with the amount of information disclosed by a person on social media. To carry out these analyses we have performed different statistical tests to find out results that provide a greater insight into online behavior pertaining to disclosure and identity.

Background & Research Questions

Social media has gained unprecedented popularity in recent years and is fast becoming the standard means of communication and maintaining relationships. Factors of mass outreach, agility, flexibility and ease of access have momentously contributed towards elevating social media from simply a means of entertainment to arguably the most efficient networking platform. However, only recently have people started to analyze into some of the online identity issues and social behavior of users on such a platform. Through this research project, we wish to analyze the afore-mentioned dataset to support our claims for the following hypotheses:

H1: Are female Twitter users less likely to use their real name (or some form of it) as Twitter usernames (primary accounts) than male Twitter users?

This research questions seeks to analyze the online identity management and the differences that exist among men and women. Pseudonyms or names that are not real are a widespread online identity behavior and it has been seen that users have been engaging in this behavior before the advent of Facebook and Twitter on older platforms such as Myspace (The Politics of "real names" 2012). In addition to this, we see that women are more likely than men to use a nickname, pseudonym, or false name online (Armentor-Cota 2011). It would be interesting to analyze this behavior for Twitter users and validate our findings.

H2: Are male Twitter users more likely to provide their location on their Twitter profiles than female Twitter users?

We see a difference in online behavior when it comes to sharing personal information among men and women. Specifically, it has been seen that men are more likely to share their personal information, such as location, than women. A survey conducted by *uSamp* on social media habits of men and women found out that men are more likely (35.3%) than women (20.1%) to share their location. It would be useful to assess whether men are more open to sharing personal information on social media platforms such as Twitter, as opposed to the opposite sex.

H3: Is there a gender-gap among Facebook users regarding the honesty and accuracy of their self-disclosures on Facebook?

Another intriguing behavior related to disclosure of information is the honesty and accuracy of information disclosed. There are significant gender differences in this behavior which suggest that women are generally less dishonest than men (Muehlheusser, G., Roider, A., & Wallmeier, N. 2015). In addition to the above, we see that women tend to justify dishonest behavior by making excuses (Ward, D. A., & Beck, W. L. 1990). It will be valuable for our discussion here to see the honesty-accuracy factor of information disclosure amongst men and women on social media and whether the same behavior applies to this domain as well or not.

H4: What are the most likely factors associated with how much a person discloses his/her true self on Facebook?

Differences in online behavior are not limited only to Gender. Each individual has a different set of principles pertaining to sharing or disclosing information about themselves. The paper titled Conceptualization and measurement of Reported Self Disclosure by Wheeless and Grotz (1976), successfully operationalizes self-disclosure as a scale consisting of multiple sub-scales such as Intended disclosure factor, Amount factor, Positive-Negative factor, Honesty-Accuracy factor, Control of General

depth factor and Relevance-message factor. This concept can be extended to self-disclosure on social media as well (Vitak J. 2012). By analyzing the data in the aforementioned dataset, we will explore the social media dynamics of users with a focus on self-disclosures generally and use of "real" identities specifically.

Method

This research project uses the dataset:
Vitak J. (2011). Facebook and Twitter Network Data. *Unpublished Raw data*. Available From 2011 Facebook and Twitter Network Data

This dataset was created by Professor Vitak by inviting, via email, a random sample of 2000 American graduate students (Master's and Ph.D.) at a large, US-based Midwestern University, to participate in an online survey about their use of online communication tools in April 2011. From this sample 486 people completed the survey. 84% of these cases (N=392) were Facebook users, while 23% of these cases were Twitter users (N=111). Twenty eight cases were deleted from Facebook user cases, due to missing data (N=364). There is only one missing data for Twitter users, and hence it was not removed. 37% of Twitter users are males and 63% of Twitter users are females. 35% of Facebook users are males and 65% of Facebook users are females.

Measures

All scale items in the dataset are measured on 5-Point Likert-type scales with options ranging from *Strongly Disagree* to *Strongly Agree*. All Frequency items are measured on a 5-point scale with options ranging from *Never* to *Very Often*.

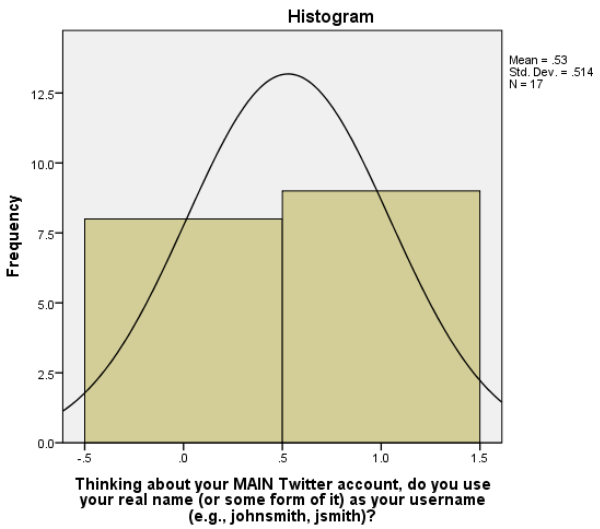
Now, we will discuss the measures or variables that will be used in the analysis of data in the following section and transformation of some of those variables and data so that they are better suited to the statistical tests that will be performed.

H1: Are female Twitter users less likely to use their real name (or some form of it) as Twitter usernames (primary accounts) than male Twitter users?

Independent Variable	**Dependent Variable**
Gender (Male:0 ; Female:1)	Do you use your real name (or some form of it) as your username (e.g., johnsmith, jsmith)?

As a starting point, we first performed a descriptive analysis of data as shown below:

Statistics		
Do you use your real name (or some form of it) as your username?		
N	Valid	17
	Missing	451
Mean		.53
Median		1.00
Std. Deviation		.514
Skewness		-.130
Std. Error of Skewness		.550
Kurtosis		-2.267
Std. Error of Kurtosis		1.063



From the Frequency analysis shown above, we can tell there are only 17 valid cases and 451 missing cases. The small number of valid cases could have been a concern for analysis, but since we are will be executing a two-sample T-Test for testing this hypothesis, we know that T-Tests are robust enough for a small sample. Also, we would not be removing any missing cases for this scenario as it would result in deletion of a major chunk of data that will be useful in testing other hypotheses.
From the histogram, we can see that the data is nearly normally distributed. The mean response for this question is 0.53. The skewness value falls in the expected range, however the kurtosis value is little higher than the acceptable range of value. But, since this variable is dichotomous, transforming the data would be futile.

H2: Are male Twitter users more likely to provide their location on their Twitter profiles than female Twitter users?

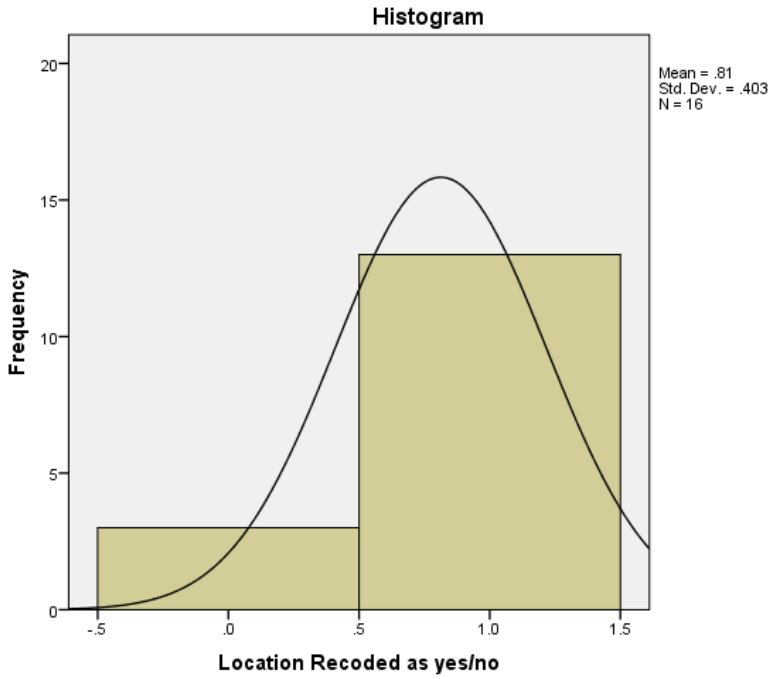
Independent Variable	**Dependent Variable**
Gender (Male:0 ; Female:1)	Main Twitter: Location

To start off, we note that the data for the Location variable needs transformation. As per the survey, it asked users if they provide any identifying information on their Twitter profile and provided check-boxes for data-entry. The data for each of those options is coded from 1-7 based on the position of the options. Therefore, all the users that checked Location were coded as 2 as per the position of Location in the list of options. However, the question was dichotomous with expected answers as “Yes” or “No”.
To achieve this, we create a new variable as Location_Recoded which consists of the codes Yes=1, No=0. All the Location values coded as 2 are treated as 1 for Location_Recoded variable. However, we find that there are only 3 Twitter users who answered the whole question and did not check “Location”. Accordingly, we assign these three data points the value 0 to signify “No” and recode 0 as 0 in the new variable.
After this processing, we perform the frequency analysis for the Location_Recoded variable as shown below:

Statistics

Location Recoded as yes/no

N	Valid	16
	Missing	452
Mean		.81
Median		1.00
Std. Deviation		.403
Skewness		-1.772
Std. Error of Skewness		.564
Kurtosis		1.285
Std. Error of Kurtosis		1.091



From the frequency analysis we can see, that again we have a very few valid cases and a lot of missing values. However, since we will be performing a two-sample T-Test it would be able to account for the small sample size. Also, we would not be removing the missing values or imputing the variables since it is a very large number.

From the histogram, we can see that the data is not exactly normally distributed. The Mean response for all the valid cases comes out to be 0.81 which tells us that more people share their location readily than those who do not. The skewness and kurtosis values are not exceedingly out of bounds and so we need not transform the data for tis variable. Also, since this variable is dichotomous, transforming the variable would not yield desired results.

H3: Is there a gender-gap among Facebook users regarding the honesty and accuracy of their self-disclosures on Facebook?

For this hypothesis, we are trying to measure the honesty-accuracy factor of people's disclosures on Facebook. The dataset consists of variables that have been categorized under different sub-scales as per the operationalization of reported self-disclosure (Wheeless and Grotz 1976). We can clearly identify the following four variables under the Honesty-Accuracy factor:

1. I always feel completely sincere when I reveal my own feelings and experiences on Facebook.
2. My self-disclosures on Facebook are completely accurate reflections of who I really am.
3. I am always honest in my self-disclosures on Facebook.
4. I do not always feel completely sincere when I reveal my own feelings, emotions, behaviors or experiences on Facebook.

Independent Variable	**Dependent Variable**
Gender (Male:0 ; Female:1	Honesty-Accuracy Scale

In order to incorporate the combined effect of all these four variables, we need to create a scale for Honesty-Accuracy factor that will accurately and completely measure this aspect. However, we can clearly see that the first three variables in our list have a positive connotation whereas the fourth variable is negatively implied. To create a robust and correct scale we need to recode the fourth variable and reverse its coding so that it gets synced with the rest of the three variables.

```
RECODE IdonotalwaysfeelcompletelysincerewhenIrevealmyownfeelingsemotion (1=5) (2=4) (3=3) (4=2) (5=1) INTO DoNoAlwaysFeelCompletelySincere_Reversed.
VARIABLE LABELS  DoNoAlwaysFeelCompletelySincere_Reversed 'Do Not Always Feel Completely Secure Variable Reversed'.
EXECUTE.
```

After recoding the fourth variable as "Do Not Always Feel Completely Secure Variable Reversed" we create an honesty-accuracy scale and conduct a reliability analysis as shown below:

Reliability

Scale: Honesty-AccuracyScale

Case Processing Summary

		N	%
Cases	Valid	373	79.7
	Excluded[a]	95	20.3
	Total	468	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.796	.797	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
I always feel completely sincere when I reveal my own feelings and experiences on Facebook.	10.6247	4.724	.702	.500	.696
My self-disclosures on Facebook are completely accurate reflections of who I really am.	10.6676	4.975	.607	.378	.745
I am always honest in my self-disclosures on Facebook.	10.3941	5.218	.611	.412	.743
Do Not Always Fell Completely Secure Variable Reversed	10.6032	5.256	.514	.283	.791

We can see from the *Reliability-Statistics* table, that the Cronbach's Alpha with all the four variables is computed as 0.796 which indicates that the scale very reliable. In addition to this, we can see from the *Item-*

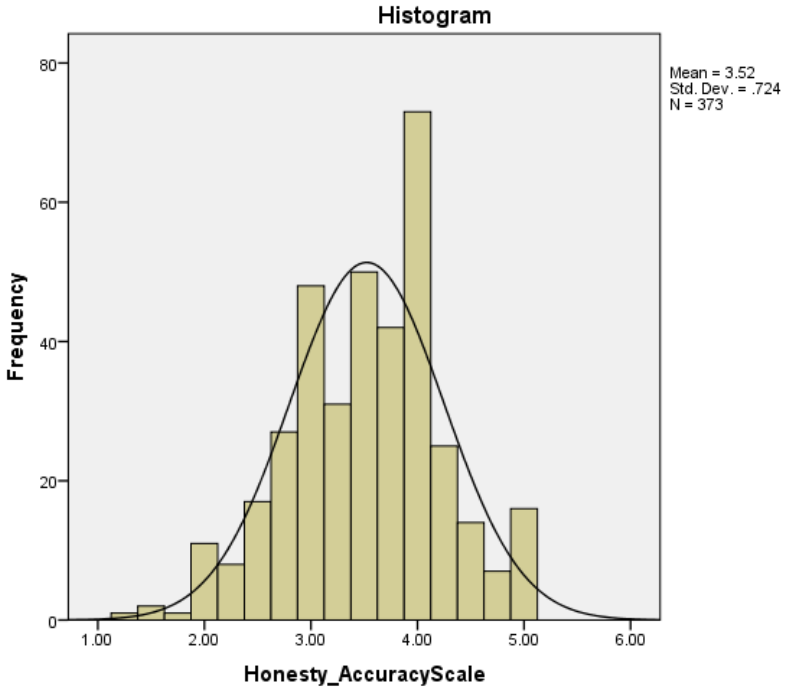
Total Statistics table that the "Cronbach's Alpha if Item Deleted" is lesser for each variable than the Cronbach's alpha of all four variables when taken together. This indicates that the scale is more reliable with all the four variables as opposed to deleting any one or more variables from the scale.

After creating the scale, we have done a Frequency analysis for the scale as well to make sure the data for the scale is normally distributed and doesn't consist of any data discrepancies.

Statistics

Honesty_AccuracyScale

N	Valid	373
	Missing	95
Mean		3.5241
Median		3.5000
Std. Deviation		.72431
Skewness		-.231
Std. Error of Skewness		.126
Kurtosis		-.066
Std. Error of Kurtosis		.252



From the frequency analysis, we can see that the data is quite normally distributed since the Mean and Median values are very close to each other and the skewness and kurtosis values are under acceptable limits as well. The *Mean* Honesty-Accuracy scale is 3.52 and it has a *Standard Deviation* of 0.72. From the histogram we can see, that the data is distributed quite normally with a symmetrical curve about the mean.

H4: What are the most likely factors associated with how much a person discloses his/her true self on Facebook?

For this hypothesis, we are trying to measure the Amount factor of people's self-disclosures on Facebook. The dataset consists of variables that have been categorized under different sub-scales as per the operationalization of reported self-disclosure (Wheeless and Grotz 1976). For this research question we have considered three sub-scales as Independent variables (IVs) and Amount factor as the Dependent variable (DV) under the self-disclosure operationalization.
We have already outlined the creation of Honesty-Accuracy scale in H3 above. For this research question we will focus on the Intended Disclosure Scale, Control of General Depth Scale and Amount Scale.

Independent Variable	Dependent Variable
Intended Disclosure Scale Honesty-Accuracy Scale Control of General Depth Scale	Amount Scale

The dataset consists of the following variables under each factor:

Scale Table

Intended Disclosure Scale:	Control of General Depth Scale:	Amount Scale:
1. When I wish, my self-disclosures on Facebook are always accurate reflections of who I really am. 2. When I express my personal feelings on Facebook, I am always aware of what I am doing and saying. 3. When I reveal my feelings about myself on Facebook, I consciously intend to do so.	1. I intimately disclose who I really am, openly and fully in my conversations on Facebook. 2. I often disclose intimate, personal things about myself without hesitation on Facebook. 3. I feel that I sometimes do not control my self-disclosure of personal or intimate things I tell about myself on Facebook.	1. I do not often talk about myself on Facebook 2. My statements of my feelings on Facebook are usually brief 3. I often discuss my feelings about myself on Facebook 4. Only infrequently do I express my personal beliefs and opinions on Facebook

We will create the scales for these factors and conduct reliability tests to ascertain their reliability and remove variables that might make the scale more reliable.

Intended Disclosure Scale:

Reliability

Scale: INTENDED DISCLOSURE SCALE

Case Processing Summary

		N	%
Cases	Valid	372	79.5
	Excluded[a]	96	20.5
	Total	468	100.0

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.696	.699	3

a. Listwise deletion based on all variables in the procedure.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
When I wish, my self-disclosures on Facebook are always accurate reflections of who I really am.	7.98	1.671	.482	.233	.643
When I express my personal feelings on Facebook, I am always aware of what I am doing and saying.	7.55	1.855	.515	.272	.607
When I reveal my feelings about myself on Facebook, I consciously intend to do so.	7.80	1.520	.547	.304	.558

We can see from the *Reliability-Statistics* table, that the Cronbach's Alpha with all the three variables is computed as 0.699 which indicates that the scale is reliable. In addition to this, we can see from the *Item-Total Statistics* table that the "Cronbach's Alpha if Item Deleted" is lesser for each variable than the Cronbach's alpha of all four variables when taken together. This indicates that the scale is more reliable with all the three variables as opposed to deleting any one or more variables from the scale.

Control of General Depth Scale:

Reliability

Case Processing Summary

		N	%
Cases	Valid	379	81.0
	Excluded[a]	89	19.0
	Total	468	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.602	.628	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
I intimately disclose who I really am, openly and fully in my conversations on Facebook.	3.39	2.117	.324	.142	.676
I often disclose intimate, personal things about myself without hesitation on Facebook.	4.10	2.320	.561	.342	.313
I feel that I sometimes do not control my self-disclosure of personal or intimate things I tell about myself on Facebook.	4.04	2.559	.392	.261	.529

We can see from the *Reliability-Statistics* table, that the Cronbach's Alpha with all the three variables is computed as 0.628 which indicates that the scale is not very reliable. In addition to this, we can see from the *Item-Total Statistics* table that the "Cronbach's Alpha if Item Deleted" is 0.676 if the first variable is deleted. This indicates that the scale is more reliable with only the second and third variable in the table. Therefore, while computing the Control of General Depth scale we do not include the first variable.

Amount Scale

For creating the Amount scale, we first need to transform three of its four variables as they have a negative connotation. The first, third and fourth variable as per the Scale table above, need to ne recoded with inverted values.

```
RECODE IdonotoftentalkaboutmyselfonFacebook (1=5) (2=4) (3=3) (4=2) (5=1) INTO
DoNoOftenTalk_Reversed.
VARIABLE LABELS  DoNoOftenTalk_Reversed 'DoNotOftenTalk Variable Reversed'.
EXECUTE.
RECODE MystatementsofmyfeelingsonFacebookareusuallybrief (1=5) (2=4) (3=3) (4=2) (5=1) INTO
StatementOfFeelings_Reversed.
VARIABLE LABELS  StatementOfFeelings_Reversed 'Statement of Feelings Variable Reversed'.
EXECUTE.
RECODE OnlyinfrequentlydoIexpressmypersonalbeliefsandopinionsonFacebook (1=5) (2=4) (3=3) (4=2)
(5=1) INTO OnlyInfrequentlyExpress_Reversed.
VARIABLE LABELS  OnlyInfrequentlyExpress_Reversed 'Only Infrequently Express variable Reversed'.
EXECUTE.
```

Reliability

Scale: AmountScale

Case Processing Summary

		N	%
Cases	Valid	379	81.0
	Excluded[a]	89	19.0
	Total	468	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.614	.594	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
DoNotOftenTalk Variable Reversed	6.4881	3.864	.482	.276	.470
Statement of Feelings Variable Reversed	7.1293	6.377	.150	.032	.675
I often discuss my feelings about myself on Facebook.	7.1108	4.564	.493	.254	.475

Only Infrequently Express variable Reversed	6.5726	3.833	.470	.227	.483

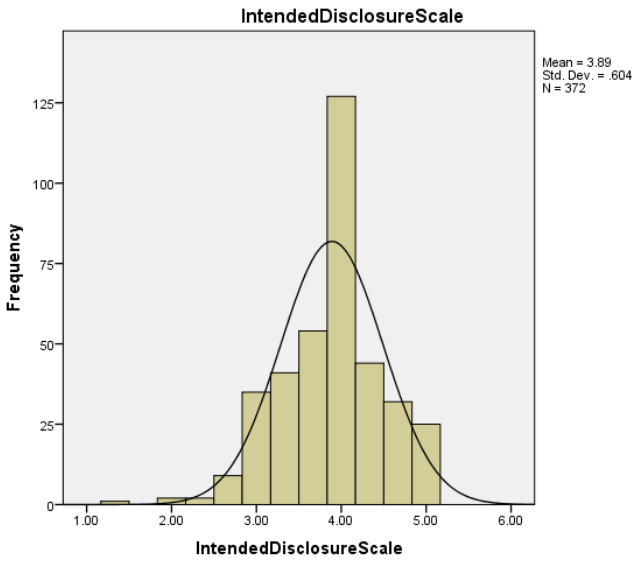
We can see from the *Reliability-Statistics* table, that the Cronbach's Alpha with all the four variables is computed as 0.594 which indicates that the scale is not very reliable. In addition to this, we can see from the *Item-Total Statistics* table that the "Cronbach's Alpha if Item Deleted" jumps to 0.67 if the second variable is deleted. This indicates that the scale is more reliable with only the first, third and fourth variable in the table. Therefore, while computing the Amount scale we do not include the second variable.

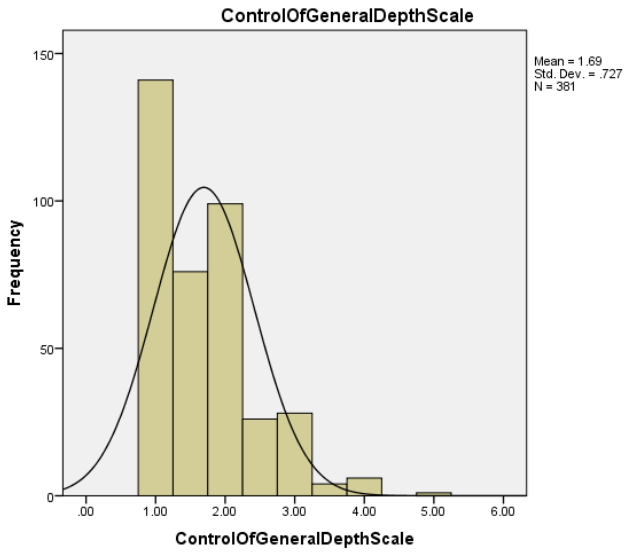
Now, we will perform a Frequency analysis to see if our scales are normally distributed or not and whether or not we need to transform their values in order to normalize the data.

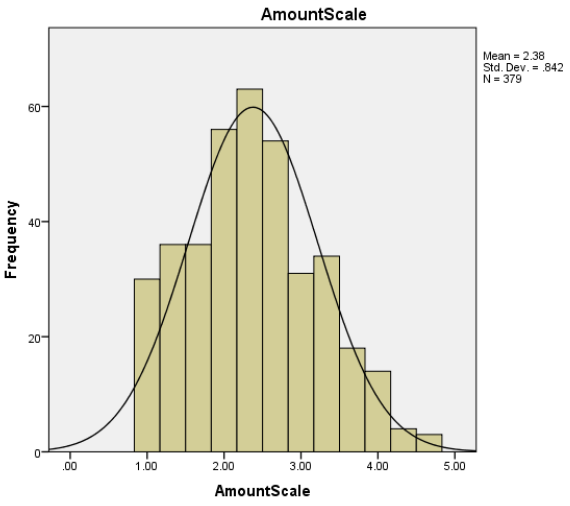
Frequencies

Statistics

		AmountScale	ControlOfGeneralDepthScale	IntendedDisclosureScale
N	Valid	379	381	372
	Missing	89	87	96
Mean		2.3764	1.6929	3.8889
Median		2.3333	1.5000	4.0000
Std. Deviation		.84179	.72651	.60414
Skewness		.304	1.139	-.405
Std. Error of Skewness		.125	.125	.126
Kurtosis		-.441	1.376	.623
Std. Error of Kurtosis		.250	.249	.252







As we can see from the Frequency descriptive table, the Amount scale and Intended Disclosure scale appear to be quite normally distributed. The Amount scale has a *Mean* value of 2.37 and a *SD* of 0.84. The Intended Disclosure scale has a *Mean* of 3.88 and *SD* of 0.6. Both of these scales have acceptable Skewness and Kurtosis values that suggest that the data is normally distributed for this scale.

The Control of General Depth scale however, has higher skewness and kurtosis values. Its Mean is 1.79 and SD is 0.72. We can transform this scale and obtain better values, but since we will be running Multiple Regression analysis for this research question, we know that this test is quite robust against data that isn't normally distributed. Also, the skewness and kurtosis values are not that alarming that it calls for transformation of data necessarily.

Analysis

In this section, we will explain how we have worked with the data and performed statistical tests for the four different hypotheses, on an individual basis.

Analysis for H1:

For the first research question, we will be performing a two-sampled T-test since our variables are dichotomous. The test results are as below:

T-Test

Group Statistics

	What is your gender?	N	Mean	Std. Deviation	Std. Error Mean
Thinking about your MAIN Twitter account, do you use your real name (or some form of it) as your username (e.g., johnsmith, jsmith)?	0	8	.50	.535	.189
	1	9	.56	.527	.176

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Thinking about your MAIN Twitter account, do you use your real name (or some form of it) as your username?	Equal variances assumed	.088	.771	-.215	15	.832		.258	-.605	.494
	Equal variances not assumed			-.215	14.712	.832	-.056	.258	-.606	.495

In the above test since the Levene's test is not significant we will consider Equal Variances are assumed, the 2-tailed p-value comes out to be 0.832. Since our hypothesis is directional, we consider 1-tailed p-value which is 0.416 and which is not significant (Significance level, α=0.05)

t(15) =- .215, p=0.832 (two-tailed) ; .416 (one-tailed) (Not Significant)

This implies that our findings are not significant and that we **fail to reject our null hypothesis**. As per our findings, there is no difference among men and women with respect to using real names as their usernames on Twitter.

Analysis for H2:

T-Test

Group Statistics					
	What is your gender?	N	Mean	Std. Deviation	Std. Error Mean
Location Recoded as yes/no	0	8	.88	.354	.125
	1	8	.75	.463	.164

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Location Recoded as yes/no	Equal variances assumed	1.577	.230	.607	14	.554	.125	.206	-.317	.567
	Equal variances not assumed			.607	13.093	.554	.125	.206	-.320	.570

In the above test since the Levene's test is not significant we will consider Equal Variances are assumed, and the 2-tailed p-value comes out to be 0.554. Since our hypothesis is directional, we consider 1-tailed p-value which come out to be 0.277 and which is not significant (Significance level, α=0.05)

t(14) = 0.607, p=0.554 (two-tailed) ; 0.277 (one-tailed) - Not Significant

This implies that our findings are not significant and that we **fail to reject our null hypothesis**. As per our findings, there is no difference among men and women with respect to sharing their location on Twitter.

Analysis for H3:

After ensuring that the data is now fit to be used in statistical analysis, we will now perform a two-sample T-Test to determine how honesty-accuracy scale differs among men and women. The results of the test are shown below:

T-Test

Group Statistics

	What is your gender?	N	Mean	Std. Deviation	Std. Error Mean
Honesty_AccuracyScale	0	129	3.4632	.73428	.06465
	1	242	3.5527	.72023	.04630

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Honesty_AccuracyScale	Equal variances assumed	.058	.811	-1.132	369	.258	-.08951	.07905	-.24495	.06594
	Equal variances not assumed			-1.126	257.051	.261	-.08951	.07952	-.24610	.06708

In the above test since the Levene's test is not significant we will consider Equal Variances are assumed, and the 2-tailed p-value comes out to be 0.258. Since our hypothesis is not directional, we consider 2-tailed p-value only as our final p-value which is not significant (Significance level, α=0.05)

t(369) =-1.132, p=0.258 (two-tailed) - Not significant

This implies that our findings are not significant and that we **fail to reject our null hypothesis**. As per our findings, there is no difference among men and women with respect to the honesty-accuracy factor of their self-disclosures.

Analysis for H4:

For this hypothesis, we are trying to measure the Amount factor of people's self-disclosures on Facebook. The dataset consists of variables that have been categorized under different sub-scales as per the operationalization of reported self-disclosure (Wheeless and Grotz 1976). For this research question we have considered three sub-scales as Independent variables (IVs) and Amount factor as the Dependent variable (DV) under the self-disclosure operationalization:
IVs:
Intended Disclosure Scale
Honesty-Accuracy Scale
Control of General Depth Scale
DV:
Amount Factor

To perform this analysis, we will be running Multiple Regression to analyze the effect of all these IVs combined on the Amount factor:

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	ControlGeneralDepthScale, IntendedDisclosureScale, Honesty_AccuracyScale[b]	.	Enter

a. Dependent Variable: AmountScale
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.430[a]	.185	.178	.76298

a. Predictors: (Constant), ControlOfGeneralDepthScale, Honesty_AccuracyScale, IntendedDisclosureScale

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	47.646	3	15.882	27.282	.000[b]
	Residual	210.154	361	.582		
	Total	257.800	364			

a. Dependent Variable: AmountScale
b. Predictors: (Constant), ControlOfGeneralDepthScale, Honesty_AccuracyScale, IntendedDisclosureScale

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.568	.312		1.817	.070		
	IntendedDisclosureScale	.207	.079	.149	2.617	.009	.695	1.439
	Honesty_AccuracyScale	.041	.066	.034	.619	.537	.728	1.373
	ControlOfGeneralDepthScale	.509	.057	.436	8.909	.000	.943	1.060

a. Dependent Variable: AmountScale

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	IntendedDisclosureScale	Honesty_AccuracyScale	ControlOfGeneralDepthScale
1	1	3.832	1.000	.00	.00	.00	.01
	2	.139	5.242	.00	.01	.02	.78
	3	.019	14.060	.22	.10	.91	.05
	4	.010	20.010	.77	.89	.06	.16

a. Dependent Variable: AmountScale

In the above test, we have used the Enter model while performing Multiple Regression and we find out that the Model is significant since its p-value is 0.00.

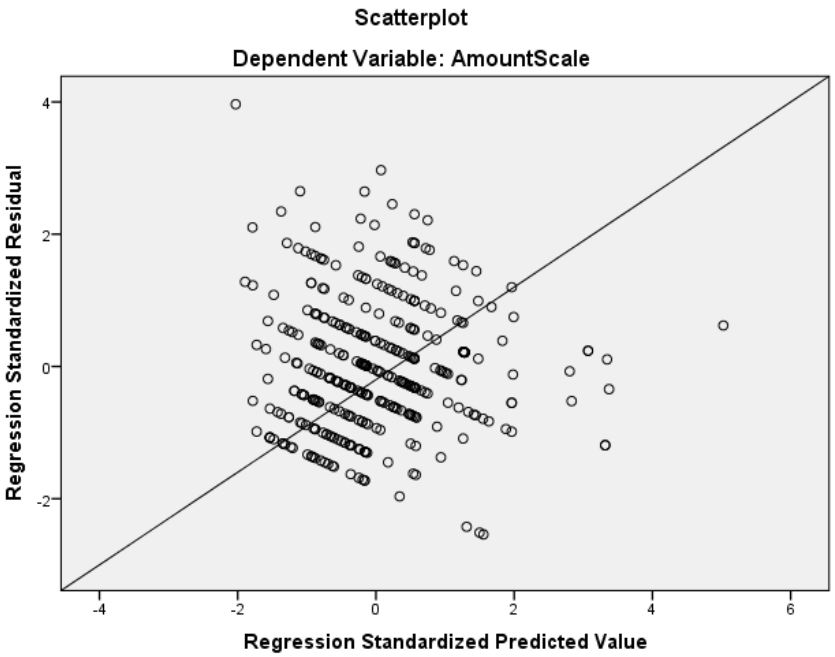
F(3)= 32.72, p=0.00 - Significant (α=0.05)

In the Coefficients table, we can also see that the Control of General Depth scale is the most significant IV of all three and Honesty-Accuracy scale is not even significant individually.

In addition to this, we see that the **Adjusted R-squared** value is 0.178 which means that the model explains for approximately 17.8% of the variance in the dependent variable.

In order to see whether are IVs have a multi-collinearity effect on regression or not, we included collinearity diagnostics as well while performing regression. In the Coefficients table, we can see that the VIF or Variance Inflation Factor for each of the IVs is low (< 3) and Tolerance is high (>0.3). This means that our **IVs do not have a multi-collinearity effect on our DV**.

Another assumption that we have tried to check for is heteroscedasticity. In order to check for this, we have created a scatter-plot of the residuals and we can see that the variance across regression line is **not heteroscedastic**, since the plot is symmetrically distributed, clustered towards the middle of the plot and clustered between 0.5 and 2 as shown below:



This implies that our findings are significant and that we **reject our null hypothesis**. As per our findings, it can be conclusively stated that Intended Disclosure factor, Honesty-Accuracy factor and Control of General Depth factors significantly predict the Amount of self-disclosure of users of Facebook.

Discussion

The research project explores the online identity and self-disclosure aspects of social media, some aspects that have been worked with only recently. These aspects provide fresh insights into the psychology of the users in the virtual world, especially on social media. The study also helps to determine if there are gender differences in some of these aspects. The study also helps us to know how much a person discloses about themselves on social media and what factors are associated with it.

From our study we have found out, that there are no gender differences in assuming online identities over Twitter. Men and women are equally likely to use their real names as usernames on Twitter. This finding is typically contrary to other existing findings. However, since this particular hypothesis focusses only on Twitter users, the result could be different from existing studies. This finding could help Twitter or other users while studying the use of pseudonyms, aiming to lessen the use or allowing pseudonyms to be used on social media and how gender profiling is not necessary in this case.

We have also found out that there are no gender differences in sharing personal information, specifically location on user's Twitter profiles. This finding is also in contradiction with exiting studies and surveys as outlined in the Background and Research questions. This could be due to the small sample size for this analysis. However, it could be a significant finding to suggest that there are no differences/diminishing differences among men and women when it comes to sharing personal information such as location on Twitter or social media in general. Such information could be valuable in extending the information disclosure behavior by Gender. Since personal information can be categorized into different sub-sections, it can be studied in greater detail in the future to determine the differences among all those categories to substantially conclude gender differences in information disclosure.

In addition to this, for our third hypothesis we found out again, that there is no gender-gap among men and regarding the honesty and accuracy of people's self-disclosures on Facebook. This finding is in contradiction with the existing findings, as per which women are less dishonest than men, in general. However, one of the existing studies was conducted in a real–world scenario such as a classroom. Our study, on the other hand focussed on honesty as well as accuracy of people's disclosures, specifically on Facebook. This helps us in concluding, that in a virtual scenario or specifically on social media, the expected behavior does not take shape. This can lead to a better understanding of users' disclosure of information about themselves for various business owners on Facebook, marketers, retailers to understand people's personalities better and create smarter solutions catering to people's preferences.

Our fourth and final finding is regarding the amount of self-disclosures of people on Facebook. Referring to the Wheeless and Grotz scale of reported self-disclosure, we analysed the prediction of Amount-factor based on the Intended-disclosure, Honesty-Accuracy and Control of General Depth factors. Our model was highly significant and revealed that the Control of General Depth factor or how much personal information a person disclosed had the most significant impact on the Amount of information disclosed. The honesty and accuracy of information revealed did not contribute individually to the amount of information disclosed. Such information can be vital to Facebook itself or its subsidiaries to target greater engagement from users. Greater engagement from users could result in greater revenues and newer markets to be explored.

To conclude, these findings shed a new light on some less-talked about and specific aspects of information disclosure on social media. Acquiring a better knowledge of the information being shared by people is vital in understanding the evolving psyche of users and creating better means of engagement for people clinging to networking in the virtual world.

References

1. Farris, D. N., Davis, M. A., & D'Lane, R. (Eds.). (2014). *Illuminating how Identities, Stereotypes and Inequalities Matter Through Gender Studies*. Springer.
2. Armentor-Cota, J. (2011). Multiple perspectives on the influence of gender in online interactions. *Sociology Compass*, *5*(1), 23-36.
3. Vitak, J. (2012). The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media*, *56*(4), 451-470.
4. Wheeless, L. R., & Grotz, J. (1976). Conceptualization and measurement of reported self-disclosure. *Human Communication Research*, *2*(4), 338-346.
5. Page, R. E. (2013). *Stories and social media: Identities and interaction*. Routledge.
6. Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters*, *128*, 25-29.
7. Ward, D. A., & Beck, W. L. (1990). Gender and Dishonesty. *Journal Of Social Psychology*, *130*(3), 333-339.
8. The Politics of "Real Names". (2012). *Communications of the ACM*, *55*(8), 29-31. doi:10.1145/2240236.2240247