When ‘friends’ collide: Social heterogeneity and user vulnerability on social network sites

Sarah L. Buglass

Jens F. Binder

Lucy R. Betts

Jean D. M. Underwood

Nottingham Trent University, UK

PRE-PUBLICATION VERSION

Published as:

BUGLASS, SARAH L., BINDER, JENS F., BETTS, LUCY R., UNDERWOOD, JEAN D.M., 2016. When ‘friends’ collide: Social heterogeneity and user vulnerability on social network sites. Computers in Human Behavior, 54, pp. 62-72.

Abstract

The present study examines how the use of social network sites (SNS) increases the potential of experiencing psychological, reputational and physical vulnerability online. From our theoretical perspective, concerns over the use of social network sites and online vulnerability stem from the ease with which users can amass large and diverse sets of online social connections and the associated maintenance costs. To date most studies of online vulnerability have relied on self-report measures, rarely combining such information with user's validated digital characteristics. Here, for a stratified sample of 177 UK-based Facebook users aged 13 to 77, digitally derived network data, coded for content and subjected to structural analysis, were integrated with self-report measures of social network heterogeneity and user vulnerability. Findings indicated a positive association between Facebook network size and online vulnerability mediated by both social diversity and structural features of the network. In particular, network clustering and the number of non-person contacts were predictive of vulnerability. Our findings support the notion that connecting to large networks of online 'friends' can lead to increasingly complex online socialising that is no longer controllable at a desirable level.

Keywords: social networks; network cluster; social spheres; network diversity; online vulnerability; online risk.

1. Introduction

1.1 Online Vulnerability and Social Network Sites

In an increasingly connected world, online social network sites (SNS; boyd & Ellison, 2008) provide interactive platforms for the digitally enabled to develop and manage their social spheres online. Surpassing the predominantly text-based methods of early computer-mediated communication, these sites afford users the ability to share a vast array of information in multimedia-rich environments. For the millions of global users who regularly engage with these sites (Ofcom, 2014), it has been suggested that they provide an online equivalent to face-to-face communication contexts (Underwood, Kerlin & Farrington-Flint, 2011), and in doing so carry the potential of delivering a range of social and psychological benefits (Burke & Kraut, 2014; Ellison, Steinfield & Lampe, 2007; Valkenburg, Peter & Schouten, 2006). At the same time, an area of mounting academic interest is addressing the potential associated risks and vulnerabilities of using SNS to interact and communicate with our social connections (Debatin, Lovejoy, Horn, & Hughes, 2009; Fogel & Nehmad, 2009; Wilcox & Stephen, 2012).

Online vulnerability is the capacity to experience detriments to psychological, reputational or physical wellbeing (Davidson & Martellozzo, 2012) due to risks encountered whilst engaging in online activities. Online risks can take on many forms (Hasebrink, Görzig, Haddon, Kalmus, & Livingstone, 2011) including threats to data privacy, online gossip and rumours, incidents of online harassment such as cyber stalking and exposure to inappropriate and unwanted content (boyd & Ellison, 2008). Recent and substantial increases in the prevalence of such adverse online experiences (BBC News, 2015; Jones, Mitchell & Finkelhor, 2013) have been linked to detrimental consequences such as depression (Landoll, La Greca, Lai, Chan & Herge, 2015) and suicide (Hinduja and Patchin, 2010, Washington Post, 2013).

Studies which have sought to find associations between SNS use and online vulnerability have so far relied mostly on self-report measures (Binder, Howes & Smart, 2012; Fogel & Nehmad, 2009;). Technological advances in data collection methods (Hogan, 2008; Rieder, 2013) now render it possible for psychologists and other researchers in non-technical disciplines to combine such information with a user's actual digital characteristics. Recently, technology-derived online network data have been used to explore social support mechanisms (Brooks, Hogan, Ellison, Lampe & Vitak, 2014). The present study will look at how such data can provide an in-depth exploration of online vulnerability that goes beyond the readily available metrics of traditional psychological research.

1.2 Online Vulnerability and Network Heterogeneity

Online vulnerability on SNS has been a frequent source of debate in both the realms of academia (Staksrud, Olafsson & Livingstone, 2013; Wilcox & Steven, 2012) and the popular press (BBC News, 2015; New York Times, 2014). Increased SNS engagement has been seen to lead to increases in online social network size (Madden, Lenhart, Cortesi, Gasser, Duggan, Smith et al., 2013), raising concerns about the consequences of network diversity (Manago, Taylor & Greenfield, 2012) and about data privacy (Debatin et al., 2009). In the following, we will outline a set of processes that link both network size and diversity to vulnerability.

SNS are typically comprised of a myriad of interconnected ego-networks (Hogan, 2008). An ego-network is a personal network in which an individual, the ego, connects with other people (Arnaboldi, Guazzini & Passerella, 2013) via a process of online 'friending'. This concept of 'friending' plays on the traditional associations conjured up by offline friendship, mutual trust, common interests and an investment of time (Thelwall, 2008), in order to encourage users to enter into a mutually agreeable digital 'friendship'. Research has suggested that many of the online 'friends' made by an ego follow an offline to online trajectory (Bryant,

Sanders-Jackson & Smallwood, 2006; Ellison et al, 2007). For the average user, SNS are an important means of maintaining pre-existing relationships (Ellison et al., 2007). This affords the ego validation and reassurance that the 'friends' viewing their data are known and trusted contacts. However, this alone may not necessarily be sufficient to guard against online vulnerability.

According to Dunbar's (1998) Social Brain Hypothesis our limited cognitive capacities and the maintenance demands exerted by social relationships impose evolutionary constraints on the size of social networks. As a result, an ego should be best equipped to maintain approximately 150 meaningful connections, i.e., contacts that have some direct relationship with ego and are characterised for the network owner by name, face, and individuating background information. Sociological studies have put the total number of people actively known to an individual, leaving aside meaningfulness, at less than 300 (McCarty, Killworth, Bernard, Johnsen, & Shelley, 2001). In the realms of SNS, however, networks regularly number in their hundreds and even thousands. Recent estimates suggest that the average adult Facebook network contains 338 'friends' (Pew Research, 2014). Whilst large networks have been positively associated with social support and informational resources (Ellison et al., 2007), a potential consequence is that they can become progressively unmanageable. One reason is that with increased size the traffic, or flow of information, through a network is likely to increase. Some proportion of this traffic will be difficult to manage for the ego (consider, for example, inappropriate broadcasting) and this proportion will likewise increase with size. Another reason is that the network's social diversity in itself becomes more difficult to manage because the ego connects to 'friends' from an increasing number of partially incompatible social spheres (Binder et al., 2012).

Each individual is highly likely to belong to a number of different social spheres and these will show up in every egocentric network. From family to friends, classmates to work

colleagues, different contacts play different roles and occupy different facets within the ego's social network. As such a social network often affords a complex structure containing multiple contextual social boundaries. In the offline world, these relationships are carefully managed by the ego enabling them to project desired and moderated representations of the self (Vitak, 2012). On SNS, however, these contextually diverse 'friends' are allowed to digitally mingle. The contextual boundaries of the heterogeneous social spheres in which they reside are collapsed, forming an increasingly homogenous online existence in the ego's network (Binder et al., 2012; Davis & Jurgenson, 2014; Marwick & boyd, 2011).

This digital mingling can lead to online vulnerability due to unintended collisions between heterogeneous social spheres. Binder and colleagues (2012), in a study on UK-based Facebook users, found that social diversity in a Facebook network resulted in increases in online tension over and above the effects of network size. This was attributed to the unrestricted flow of information across the collapsed contextual social boundaries. For example, a 'friend' of the ego posting information pertinent to the sphere in which they reside (e.g. a risqué 'in' joke) might inadvertently cause tension with 'friends' from contextually different spheres within the network.

In a contextually collapsed network, however, it is not just the risk posed by the communications of the ego's friends that can potentially increase vulnerability, but also the communications of the ego themselves. SNS impact on our ability to imagine the audience to which we are communicating (boyd, 2007; Litt, 2012). When we engage in communication with individuals or small groups (i.e. in face-to-face settings or via small scale technology-mediated communications), the audience to whom we are communicating is unambiguous due to immediate visual and/or auditory validation (Litt, 2012). On social networking platforms, however, audiences have a tendency to become less explicit as the size, diversity and permanence of the networks increasingly decreases their salience (boyd, 2007).

When an ego posts a communication on an SNS, it is likely that their imagined audience does not consist of the complete social network but rather a subset derived from either technological cues (e.g. the 'Online' friend list, frequent likers/commenters) or cognitive references to offline social contexts (Marwick and boyd, 2011). For the ego, this potential to misjudge the prospective audience has implications for online vulnerability, due to an increased likelihood in the ego communicating content that is not appropriate for all of the heterogeneous social spheres contained on their network (Binder et al., 2012). On this basis, we expected first of all that network size and social heterogeneity would both be positively related to vulnerability:

H1: Network size will positively predict exposure to online vulnerability.

H2: Social network heterogeneity will positively predict exposure to online vulnerability.

Heterogeneous spheres so far have been defined and measured as social diversity, the different types of contacts that can be identified in a network (Binder et al., 2012; McCarty et al., 2001). This leaves the question how these contacts are arranged and interconnected. SNS carry the unique advantage of digitally mapping out network structures, which allows for the identification and quantification of clusters (Smith, Schneiderman, Milic-Frayling, Mendes Rodrigues, Barash, Dunne et al., 2009). Clusters are discernible subgroups characterised by a high degree of internal interconnections and few external connections to other parts of the network. As such, they provide another indicator of different spheres managed by ego. Clusters may not fully coincide with the social categories listed for a network. For example, a category 'friends known from school' may be located within one cluster representing the social environment of ego at school and another cluster representing an inner friendship circle that is distinct from the wider school context. In this study, we considered not only the diversity of

social contacts as identified by ego but also the actual heterogeneous clustering of ego's online network. We thus hypothesise that:

> H3: Structural network heterogeneity will positively predict exposure to online vulnerability.

In addition, we tested a more comprehensive model to integrate network size, heterogeneity and vulnerability. While previous research has shown that heterogeneity can have effects independent of size (Binder et al., 2012), findings also suggest that problematic online incidents may well be related to network size through an increase in heterogeneity (Manago et al., 2012). In other words, network size is a driver for developing those network characteristics that lead to higher levels of online vulnerability, and the size-vulnerability relationship is mediated by these characteristics. We therefore propose that:

> H4: Social and structural heterogeneity will mediate the relationship between network size and online vulnerability.

1.3 Implications of non-standard online 'friends'

Ego-centred sites such as Facebook and LinkedIn actively encourage people to provide a wealth of personal information to aid validation of user authenticity and guard against instances of fake profiles. While some studies have shown online presentations of the self to be are generally accurate (Back, Stopfer, Vazire, Gaddis, Schmukle, Egloff et al., 2010; YouYou, Kosinski, & Stilwell, 2015), it has been estimated that approximately 5 to 11 percent of Facebook profiles might be erroneous (Facebook, 2015).

Safely navigating an online network may also be compromised by the presence of 'friends' who are not characteristic of traditional online connections. Most SNS, and indeed most Internet services, do not recognise individuals, but user accounts. The assumption, however, that all user accounts represent true, individual people is not warranted. Accounts

may also include or omit information that is important for ego to reliably identify other contacts. Non-standard online contacts can therefore make it even more difficult for a user to form an impression of their actual audience.

At present it is not possible to identify with great certainty profiles on a network that might offer negative consequences to the ego and their connections. However, digital ego network data offer some opportunities to identify characteristics indicative of 'non-standard' connections. Here, we are particularly concerned with misclassified profiles, use of obvious pseudonyms, missing information and socially isolated contacts.

1.3.1 Misclassified profiles

Misclassified profiles occur when the social network account holder creates a profile that does not match the general norms or expectations of a traditional profile. According to Facebook's (2015) annual report to the US Securities Exchange Commission, approximately 2% of all monthly active profiles on Facebook are misclassified profiles. Whilst 2% may not at first appear substantial, in the context of Facebook which currently has approximately 1.39 billion monthly active users, this equates to an estimated of 27.8 million profiles.

Misclassified profiles are entities that should be represented on an online social network by a 'page' or specific space and not by a personal profile. They are often representative of small companies, organisations, social interest groups and even pets. Misclassified profiles may occur due to user-error (i.e. the account holder is not familiar with the terms and conditions of the site) or potentially for malicious purposes (i.e. a person pretending to be a known company using a fake profile in order to gain data/and or money from unsuspecting users).

1.3.2 Pseudonym use

The use of a pseudonym is a form of identity concealment (Hogan, 2012). Full pseudonyms offer a completely non-representative name – often made up or indicative of a figure from

popular culture. Partial pseudonyms might use one of the individual's real names in addition to a "made up" name, i.e. Super Sarah. A number of high profile SNS implement a 'Real Name Policy' for which they actively encourage the use of real names (Facebook[2], 2015; LinkedIn, 2015). The policy is indicative of a growing trend on online platforms toward non-anonymised communication (Hogan, 2012), driven in part by a desire to impinge on the growing problem of fake or erroneous profiles. Whilst the presence of pseudonym profiles on the network is not necessarily indicative of potential harm to the ego (Hogan, 2012) it has been suggested that such online anonymity may increase the likelihood of anti-normative behaviour being experienced (Cho, Kim & Acquisti, 2012).

1.3.3 Inaccurate or missing data

Inaccurate or missing data in profiles does not match the general norms or expectations of a standard social networking profile. As suggested by Herring and Martinson (2004), the non-disclosure of personal attributes, such as gender, not only potentially impedes an ego's ability to authenticate the identity of their prospective connection, but may also limit opportunities for them to moderate their communications in a manner appropriate to the norms and conventions associated with their prospective connections.

1.3.4 Social Outliers

Social outliers are individuals that are connected to the ego only. They are socially distant contacts who do not share any mutual friends with the ego and as such lack validation from other members of the ego network. Whilst some have theorised that such bridging or weak ties can provide the ego with diversified social and informational support (Burt, 2000), others have suggested that outliers may promote friction within the network as they have the fewer social and reputational costs (Brass, Butterfield & Skaggs, 1998). Interestingly, outliers may in time become more highly connected within the network. Boshmaf et al. (2011), for example, found

that SNS users were almost 50% more likely to accept a friend request if the connection had at least one mutual friend.

The presence of non-standard network connections has the potential to further complicate the ego's ability to effectively manage and moderate their online communications. While users view their close social spheres as points of reference for generating their target audience on social media (Marwick and boyd, 2011), sporadic cases of non-normative profiles are likely to be less salient. A potential consequence of this lack of salience is further social tension due to contextual collapse: from the perspective of both the ego and the non-standard profile holder. Additionally, ego's vulnerability to malicious behaviours such as identity theft, data misuse and harassment is likely to increase due to the privacy implications of sharing data and communications with profiles that cannot be readily authenticated. In sum, we expected all non-standard connections identified to contribute to online vulnerability. Thus we hypothesised that:

> H5: The presence of Facebook profiles demonstrating non-norm characteristics will positively predict exposure to online vulnerability.

Research has linked increases in network size and diversity to increases in superficial and unknown contacts (Manago et al., 2012). Assuming a small percentage of non-norm characteristics to be present in most active Facebook networks, it follows that the absolute frequency of such characteristics will increase with growing network size. Networks that run into hundreds, or thousands, of online contacts are no exception on Facebook and are likely to exhibit a non-negligible number of non-norm characteristics for mere probabilistic reasons. Furthermore, studies have also suggested that users holding larger networks may be more inclined to engage in "promiscuous friending activities" (Stefanone, Lackaff & Rosen, 2011; Stefanone, Lackaff & Rosen, 2008). From this perspective, the more the ego engages in these

activities, the less consideration the ego will give to a profiles actual validity or status, when adding online contacts. Therefore, we also expect that:

H6: Frequency of non-norm network characteristics will mediate the relationship between network size, diversity and online vulnerability.

2. Method

An integrated data set was generated from cross-sectional survey measures and digitally derived network data to explore the relationship between Facebook network characteristics and online vulnerability.

2.1 Sample

Self-report survey data and digitally derived Facebook metrics were obtained from an opportunity sample of 177 UK based Facebook users (63% female). Participants were recruited from three UK-based populations stratified by age:

(1) Secondary school aged children (N=50) between 13 and 17 years from three socio-economically diverse UK schools. School and parental consent were obtained prior to the study.
(2) Undergraduate students (N=63) from a large UK university. Participants responded to advertisements placed on student bulletin boards and also via a departmental participant pool. Research credits were awarded for participation in the study.
(3) Online adult users (N=64) recruited via online advertisements. Permissions were gained from the administrators of the online message boards and communities prior to any advertisements being displayed.

In return for their time, all participants were eligible for entry into a prize draw to win online vouchers. Appropriate ethical procedures were observed for all three sub samples.

2.2 Measures

2.2.1 Self-reported measures

Study-specific self-report measures were used to determine user and Facebook demographics, rate of exposure to online vulnerability and the number of different social 'friend' types connected to via Facebook.

User Demographics. Items addressing age and gender (coded as 0 for male, 1 for female) were measured in order to provide a general overview of the sample characteristics.

Facebook Demographics. Addressed by three items: duration of Facebook membership (in years); rate of daily Facebook engagement (up to 15/30/45/60 min; more than 60 min) and current Facebook privacy settings (e.g. "Friends Only").

Online vulnerability. Assessed using a six item scale adapted from questions and theory presented in Binder et al. (2012). Participants were asked to indicate how frequently they had been exposed to a range of online vulnerabilities (e.g. "critical or hurtful comments", "social embarrassment", "damaging gossip and rumours", "content of sexual or violent nature", "unwanted attention" and "data misuse") whilst using Facebook. Responses to each item were positively anchored and ranged from 1 (Very Rarely) to 5 (Very Often). Items were averaged to form a reliable index (α=.93) with higher values indicating increased exposure to online vulnerability.

Social 'Friend' Types: Sixteen friend types, listed in Table 3, were presented as dichotomous (Yes/No) items. The items were adapted from common network cluster categories previously attributed to ego-centric social network structures (Binder et al., 2012; McCarty et al., 2001). An overall tally of the number of different friend types was produced by summing up the number of positive responses to these items. Scores could therefore range from 0 to 16, with higher scores indicating increased heterogeneity of connections in the social network.

2.2.2 Digitally derived network characteristics

Network characteristics were derived from data generated by Netvizz (Rieder, 2013), an application that enables individual Facebook users to access their mutual friendship data generated by the Facebook API (application programmer interface). Network data obtained this way include a unique identifier for each Facebook contact, the name of the Facebook contact and their gender. Further, all interconnections among ego's contacts are listed. Facebook users who have set high privacy permissions are not captured by the application. For this reason Netvizz data can only provide an estimate of the actual structural properties of the user's network. Network metrics were calculated using NodeXL, a network analysis tool developed by the Social Network Research Group (Hansen, Shneiderman & Smith, 2011).

Network Size. An estimate of digitally derived network size was gained by summing the total number of network contacts listed in the Netvizz data. Network sizes for this sample ranged from 4 to 1468.

Network Clustering. Clustering was calculated using the Clauset-Newman-Moore (2004) algorithm. A clustering coefficient was created for each individual node within the network. A global clustering coefficient was then produced for the entire network by averaging the individual coefficients. The global clustering coefficient ranges from 0 to 1. As exemplified in Figure 1, coefficients approaching 1 indicate closely knit networks with dense network structures with only a small number of social spheres present in the network. In contrast, coefficients closer to zero, as exemplified by Figure 2, are indicative of more heterogeneous network structures encapsulating multiple social spheres, isolated connections and instances of anomalous network contacts.

Network Anomalies: profiles that are not characteristic of personal profile norms and/or patterns of connectivity evident in typical Facebook networks. Anomalies were measured by

four variables: gender-hidden profiles, misclassified profiles, pseudonym represented profiles and network outliers. *Gender-hidden profiles* were calculated using gender information for each network contact derived from the digital data. The number of network contacts with missing gender details was summed. This provided a total score of gender-hidden network contacts for each individual network. The total number of *network outliers* was generated using social network analysis to identify the number of network isolates in each individual network.

To calculate the number of *misclassified profiles* and *pseudonym represented profiles*, a qualitative appraisal of the network contacts was made. All network contacts were inspected across the 177 networks (approximately 71,000) for instances of obvious pseudonyms (e.g. Mickey Mouse) and/or misclassified entities (e.g. companies, student groups) using a study-specific set of anomaly indicators. This was done by one rater. A sample of 1,500 network contacts was then given to a second rater and ratings were compared. Where raters disagreed this was resolved without difficulty indicating good general understanding of the coding criteria. Further, Cohen's κ showed good inter-rater agreement (κ = .73 (95% CI, .67 to .80), $p < .001$). Instances of pseudonyms and misclassified entities were then summed up to provide an overall total for each network.

2.3 Procedure

Participants completed a secure online survey, optimised for use on desktop computers, tablets and mobile devices. On completion of the self-report survey participants were asked to obtain and send through their digitally derived network data. This part of the procedure was fully integrated with the survey.

3. Results

3.1 Preliminary Analyses

General sample characteristics are displayed in Table 1. The mean age of the sample was 22 years 10 months (*SD* = 9.82; *Range*: 13-77 years). The mean duration of Facebook membership was 5 years 5 months (*SD* = 2.04 years). Over half of all participants (54%) reported engaging with Facebook for 30 minutes or less per day. However, the majority of participants (72%) reported high rates of actual connectivity, indicating that whilst not actively engaging with Facebook they very rarely logged out of the network. The majority of participants (89%) reported using at least the standard "Friends Only" Facebook privacy settings, with 22% of these using more advanced additional filtering options.

Descriptive statistics for the main measures are given in Table 2. Participants had on average experienced a moderate level of overall online vulnerability whilst using Facebook (*M* = 2.75, *SD* = 1.09, on a scale from 1 to 5). Network variables, given their scale, were not normally distributed, which was taken into account in subsequent analyses. Network size, for example, had a mean of *M* = 399.40 and a range of 4 – 1468 (*SD* = 277.25). The presence of a small number of large networks containing over 1000 friends led to a positive skew.

A closer inspection of the types of friends listed (see Table 3) indicated that friends/class mates and family members were most frequent among network contacts. However, it should be noted that 62% of respondents named casual acquaintances, 28% online only contacts and 25% public figures among their contacts.

In order to control for the non-normal distribution of the network derived data Spearman's Rho correlation coefficients were calculated. These indicated the association between online vulnerability and the different measures of social network characteristics (see Table 4). The correlations amongst the main study variables did not suggest multi-collinearity with only one coefficient > |.07|.

As expected digitally derived network size moderately correlated with the measure of online vulnerability (r_s = .38, p < .001), indicating that as network size increased online vulnerability scores also increased. Furthermore, as network size increased, the number of social friend types (r_s = .43, p < .001) and the rate of clustering also increased (r_s = -.51, p < .001), indicating increased level of both social and structural heterogeneity. Network size was also positively correlated with instances of anomalous network contacts.

Online vulnerability moderately correlated with social friend types (r_s = .37, p < .001) and network clustering (r_s = -.26, p < .001), with scores in online vulnerability increasing as social and structural network heterogeneity increased. Correlations between online vulnerability and the anomalous network contacts were more mixed. Increases in misclassified profiles (r_s = .39, p < .001), pseudonyms (r_s = .20, p < .05) and network outliers (r_s = .17, p < .05) were associated with increases in online vulnerability. No significant association was found between gender-hidden profiles and online vulnerability. All network anomalies were significantly correlated with both network clustering and social 'friend' types, with the only exception being the relationship between social 'friend' types and gender-hidden profiles (*ns*).

3.2 Structural and Social Predictors of Online Vulnerability

In order to test H1, H2 and H3, a set of bootstrapped hierarchical regression analyses were performed with online vulnerability as the dependent variable. Due to initial violations of normality and linearity, all variables were square root transformed prior to the analysis[1]. Following the transformation all assumptions of multiple regression were met. An overview of the regression analyses can be found in Table 5.

H1 stated that network size would be positively related to online vulnerability. In the first instance digitally derived Facebook network size was entered as the independent variable. Participant age and gender were controlled for. The overall regression model was

significant (F (3,176) = 12.43, p < .001), accounting for 17.7% of the variance of online vulnerability. In line with the initial correlational analysis, an increase in network size was predictive of increases in online vulnerability (b = .02, β = .34, p < .05) thus confirming H1. In addition, age was negatively related to vulnerability (b = -.05, β = -.13, p < .05).

H2 and H3 stated that network heterogeneity would be positively related to online vulnerability. The regression model was therefore expanded to include social 'friend' types and network clustering as predictors of online vulnerability. Once again the overall model was significant (F (5, 176) = 10.73, p < .001), now accounting for 23.9% of the variance of online vulnerability. This represented a significant 6.2% change in the R^2 value from the previous model (p = .001).

Network clustering and reported social 'friend' types added statistically significantly to the predictive model (p < .05). The standardised beta coefficients indicated that increases in network diversity, as typified by increases in the number of social 'friend' types (b = 1.32, β = .18, p < .05) and decreases in the network clustering coefficient (b = -1.23, β =-.21, p < .05), are predictive of increases in network size. This means both H2 and H3 were supported. Again, age was a significant and negative predictor in the model (b = -.07, β = -.19 p < .05) suggesting that online vulnerability might be more apparent in the younger Facebook users amongst the sample. Introducing social groups (friend types) and network clustering to the model rendered the predictive value of network size insignificant. This was indicative of a potential mediating influence of these variables on the relationship between network size and online vulnerability, tested in detail further below.

The presence of network anomalies, as postulated by H5, should predict online vulnerability. To test this, the final regression model added number of misclassified profiles, gender-hidden profiles, pseudonym-represented profiles and network outliers to the predictors.

The addition of these variables imposed a significant 4.8% change in the R^2 value (p = .03) increasing the total variance explained for online vulnerability to 28.7%. Of the four anomalies identified in the data, only misclassified profiles proved to be significant (b = .07, β = .24, p < .05). Network size and age were not significant in Model 3. In sum, H5 received partial support.

3.3 Mediating the Effects of Network Size on Online Vulnerability

In the following, we report a set of mediation analyses covering H4 and H6. We adopted a bootstrapped multiple mediation approach (Preacher & Hayes, 2008), using PROCESS (Hayes, 2015), a macro developed for use with SPSS. Such models have been likened to structural equation models in that they enable researchers to consider which part of an explanatory variable's effect on a dependent variable can be explained by a mediating variable (Brooks et al., 2014).

H4 stated that effects of network size on vulnerability would be mediated by social and structural heterogeneity. The model testing this hypothesis is illustrated in Figure 3. An analysis of the 95% BCa confidence intervals (Table 6) of the indirect effects of social 'friend' types and network clustering indicated that they significantly mediated the association between Facebook network size and online vulnerability. Both mediated paths were found to be significant in terms of both the traditional Sobel Test (p < .05), associated with the Baron and Kenny (1986) causal steps approach to mediation and also via the analysis of the bootstrapped confidence intervals generated by the indirect effects (Preacher & Hayes, 2008). Furthermore, the completely standardised indirect effect (β = .20, 95% BCa CI [.10, .32]) was indicative of a moderate overall effect size for the model. This means that H4 received full support.

As shown in Figure 3 the indirect effect of social 'friend' types was found to have a positive association with network size (Path a: β = .03) and a positive association with online

vulnerability (Path b: β = .13). These results imply that increased network size increases the number of social 'friend' types in the network, which in turn increases the likelihood of exposure to online vulnerability. The indirect effect of network clustering was found to have a negative association with network size (Path a: β = -.01) and a negative association with online vulnerability (Path b: β = -1.23). As lower network clustering coefficients are indicative of higher network diversity it is appropriate to interpret these results in terms of increases rather than decreases. The indirect effects therefore imply that increased network size usage increases network diversity via clustering, which in turn increases the likelihood of exposure to online vulnerability.

Finally, a mediation model was tested to further investigate the hypothesised role of network anomalies in the network-vulnerability relationship (H6). The model, fully shown in Figure 4, considered potential indirect effects from both the perspective of parallel and serial mediators. The analysis of serial multiple moderation effects via the PROCESS macro does not produce an indication of significance via the traditional Sobel test. Alternatively an analysis of the 95% BCa CI bootstrapped tests is used.

The confidence intervals for the model (Table 7) indicated that there were some significant indirect effects present between the association of network size and online vulnerability. The overall model produced a completely standardised indirect effect (β = .29 95% BCa CI [.16, .44) which was indicative of a moderate overall effect size for the model.

In terms of the parallel indirect effects, social 'friend' types continued to be a significant mediator in the relationship between network size and vulnerability, indicating as before that increases in network size increase the number of social 'friend' types in the network, which in turn increase the likelihood of the ego being exposed to online vulnerability. Misclassified

profiles also offered a significant indirect effect, with increases in network size leading to increases in misclassified profiles, and in turn an increase in online vulnerability.

Interestingly, the inclusion of misclassified profiles appeared to render the indirect path relationship between network clustering and online vulnerability (Figure 4) non-significant. However, the overall indirect effect between network size, network clustering and online vulnerability appeared to remain significant in terms of the overall bootstrapped indirect effect (Table 7), although the overall effect size was somewhat diminished. This result complimented the previous findings of the hierarchical regression analysis in being suggestive of non-person profiles playing a potentially mediating role in this relationship.

The mediating role of non-person profiles on the relationship between network clustering and online vulnerability was confirmed via the analysis of the serial indirect effects in the model. A significant serial indirect effect was found between network size, network clustering, misclassified profiles and online vulnerability. This significant effect was evident in both the path relationships (Figure 4) and also the overall bootstrapped effect (Table 7). Allowing for the backwards interpretation of network clustering, the indirect effect implies that increases in network size lead to an increase in network diversity (due to a decrease in the network clustering coefficient). Increases in network diversity then lead to increases in misclassified profiles, which in turn result in increased likelihood of the ego being exposed to online vulnerability.

Non-person profiles were not found to have a significant indirect effect on the relationship between social ‘friend’ types and online vulnerability. However, when social ‘friend’ types was considered as a serial mediator with both network clustering and non-person profiles it did produce significant indirect effects on the relationship between network size and online vulnerability. As such increases in network size appeared to increase the number of

social ‘friend’ types. Increases in social ‘friend’ types leads to reductions in the network clustering coefficient, therefore increasing network diversity. Increases in network diversity increase the likelihood of non-person profiles being present in the ego network, which in turn increases the likelihood of the ego being exposed to online vulnerability.

4. Discussion

The present study explored the impact of social and structural network characteristics on the online vulnerability of ego-centric social network site users. Utilising a mixed methods approach to online data collection and analysis, the results provide an innovative examination of online social networking characteristics. The main findings can be summarised as follows. First, consistent with the network size hypothesis (H1), increases in network size were indicative of increases in online vulnerability. Second, consistent with the hypotheses that social and structural heterogeneity positively predicts online vulnerability (H2 and H3); increases in self-reported social diversity and digitally derived network diversity were predictive of increases in online vulnerability. Furthermore, social and network diversity mediated the relationship between network size and online vulnerability (H4). Third, partial support was obtained for the network anomalies hypothesis (H5), that profiles exhibiting non-standard network characteristics are positively predictive of online vulnerability. Misclassified profiles were predictive of increases in online vulnerability, no other non-standard characteristics were found to be significant predictors. Misclassified profiles also provided a mediating role in the relationship between structural characteristics and online vulnerability (H6).

4.1 Network size and social heterogeneity

The findings revealed that individuals with larger network sizes tended to be more prone to experiencing online vulnerability on ego-centric social networking sites, largely due to

increases in the social and structural diversity of their networks. One explanation for this is contextual collapse. As the number and variety of online contacts increases, the boundaries between heterogeneous social spheres collapse (Vitak, 2012), rendering it difficult for the ego and their contacts to effectively imagine their target audience when sharing content (Litt, 2012; Marwick & boyd, 2011). Content intended for a particular 'imagined' sphere becomes visible across the network, often with little regard for its appropriateness for those outside the 'imagined' sphere. The high visibility of such unmoderated content on ego-centric online networks facilitates increases in network tension (Binder et al., 2012) within the network and also potential vulnerability of the ego and their contacts, due to the increased risk of exposure to potentially contentious and inappropriate material.

A novel aspect to this perspective is provided by our finding that the different types of contacts and the clustering of these contacts were both predictive of vulnerability. Put differently, clusters did not align with categorisation of contacts, and both sources of information independently help to explain the challenges that arise from the maintenance of online networks. Social spheres as clusters may refer to particular life stages (e.g., contacts from school days) or to particular environments (e.g., contacts from the office), in which case they would still be likely to contain a range of social ties. Conversely, social spheres as different categories of others may well be distributed over several clusters (e.g., all closer friends, no matter where they are usually encountered). Broadcasting in SNS therefore jeopardises the balance within clusters as much as between clusters. Addressing the exact composition of clusters in terms of categories of others is beyond the scope of the present work, but immediately suggests itself to put our speculations here to the test.

4.2 Network Anomalies

The occurrence of non-standard network profiles rendered mixed results. Misclassified profiles were found to significantly predict increases in online vulnerability. A possible reason for this is that misclassified profiles represent a diverse array of non-personal entities. When an ego connects to a misclassified profile, they share their personal timeline and content with the likes of businesses, student/interest groups and possibly fake profiles. Many users of ego-centric online social networks knowingly upload and share vast amounts of data (Debatin et al., 2009). Misclassified profiles, therefore, gain potential access to the ego's likes, dislikes, location and photographs, presenting the ego with a potential minefield of opportunities for data driven online vulnerability such as data misuse and identity theft, which may ultimately impact on their psychological, reputational and physical wellbeing.

Interestingly misclassified profiles were also found to mediate the relationship between structural characteristics of network size and diversity and online vulnerability, indicating that increases in online vulnerability in large and structurally diverse networks are potentially enhanced by the presence of misclassified profiles. In a large, structurally diverse network, misclassified profiles may make the imagined audience unimaginable, as the ego is presented with the complex task of determining not only 'who' but 'what' they are sharing their content with.

A more unexpected result was the insignificance of the remaining non-standard profile characteristics as predictors of online vulnerability. Whilst initial correlational analysis had rendered a non-significant finding for gender concealment, small but significant positive relationships with online vulnerability for both pseudonym use and network outliers were found, indicating potential to support the network anomalies hypothesis. However, further analysis of the predictive significance of these anomalies provided inconsequential results.

In the case of pseudonym use and gender-concealment: the predictive non-significance of these non-standard characteristics calls into question a core argument of the 'real-name' policies currently being mooted by many online social networking sites (Hogan, 2012). Promoters of the policy claim that such forms of identity concealment might promote potentially negative behaviours on a network and therefore increase the online vulnerability of wider network users (Cho et al., 2012; Hogan, 2012). The results of this study imply that individuals adopting such non-standard characteristics may not necessarily be ill-intentioned and may in some cases merely exercising their right to express their identity online in a manner unbound by the potential risks and restrictions of non-anonymised data exchange.

The insignificant predictive association between network outliers and online vulnerability was also unexpected. Prior research had suggested that unconnected individuals in a network might increase tension and vulnerability due to the low social and reputational costs of their potential exchanges online (Brass et al., 1998). Whilst correlational analysis did provide minor evidence for this theoretical standpoint, the lack of predictive significance suggested that network outliers might not necessarily constitute online vulnerability in all networks.

The mixed results rendered by the network anomalies indicate that further research is required in order to determine the role that such entities have within an ego-centric online social network. Ego-centric online social networks have amassed global participation numbering in the billions. In contrast, the present study provides a cross-sectional snapshot of only 177 users. With this in mind further large scale, longitudinal analysis is recommended.

To conclude, this present study provides significant support for the relationship between social and structural network characteristics and online vulnerability. In doing so, it increases our understanding of the potential detrimental effects of the contextual collapse of

social spheres on online networks by adding digitally derived information to the largely self-report based theoretical standpoints of previous social network literature (Binder et al., 2012; Vitak, 2012). Furthermore, the study provides an indication of the potential for vulnerability that may be brought via connecting to certain types of anomalous and potentially nefarious network contacts. These findings carry implications for those designing SNS and applications integrated with SNS technology. To the extent that an automated recognition of types of contacts can be improved (Eagle, Pentland, & Lazer, 2009) and pseudonyms can be more reliably detected, a combination of indicators of social heterogeneity and non-standard profiles could be used to identify vulnerable users with the aim of offering them software settings and advice to better protect them. Closely related to such interventions is the emergent and pervasive problem of maintaining any substantial level of data privacy on social media. As a continuous flow of news headlines suggest, the breakdown of online privacy has, in some cases, severe implication for individual wellbeing. In a world that experiences a marked shift towards everyday online experiences and the use of virtual social spaces, experts and the general public need models that guide us in making these experiences and spaces psychologically safe and sound.

Footnotes

[1] Untransformed regressions were run as a comparison. Comparable findings for the first two models were evident, although the magnitude of effect sizes was slightly reduced. For model 3 the introduction of the heavily skewed pseudonym and outlier data had a substantial impact on the results overall. This provided support for the use of transformed data in the models.

References

Arnaboldi, V., Guazzini, A., & Passarella, A. (2013). Egocentric online social networks: Analysis of key features and prediction of tie strength in Facebook. *Computer Communications, 36(10),* 1130-1144. doi:10.1016/j.comcom.2013.03.003.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*. doi: 10.1177/0956797609360756.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51(6)*, 1173. doi: 10.1037/0022-3514.51.6.1173.

BBC Online, 2015, "Who's that girl? The curious case of Leah Palmer." Available at: http://www.bbc.co.uk/news/technology-31710738 Accessed on: 24/04/2015.

Binder, J. F., Howes, A., & Smart, D. (2012). Harmony and tension on social network sites: Side-effects of increasing online interconnectivity. *Information, Communication & Society, 15(9),* 1279-1297. doi: 10.1080/1369118x.2011.648949.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011, December). The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference* (pp. 93-102). ACM. doi: 10.1145/2076732.2076746.

boyd, D. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. *MacArthur foundation series on digital learning – Youth, identity, and digital media volume*, 119-142.

boyd, D., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13(1),* 210-230. doi: 10.1111/j.1083-6101.2007.00393.x.

Brass, D. J., Butterfield, K. D., & Skaggs, B. C. (1998). Relationships and unethical behavior: A social network perspective. *Academy of Management Review, 23(1),* 14-31. doi: 10.5465/amr.1998.192955.

Brooks, B., Hogan, B., Ellison, N., Lampe, C., & Vitak, J. (2014). Assessing structural correlates to social capital in Facebook ego networks. *Social Networks, 38,* 1-15. doi: 10.1016/j.socnet.2014.01.002

Bryant, A. J., Sanders-Jackson, A., & Smallwood, A. M. (2006). IMing, text messaging, and adolescent social networks. *Journal of Computer‐Mediated Communication, 11(2),* 577-592. doi: 10.1111/j.1083-6101.2006.00028.x

Burke, M., & Kraut, R. E. (2014, April). Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the 32nd annual ACM Conference on Human Factors in Computing Systems* (pp. 4187-4196). ACM. doi: 10.1145/2556288.2557094.

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior, 22,* 345-423. doi: 10.1016/s0191-3085(00)22009-1

Cho, D., Kim, S., & Acquisti, A. (2012). Empirical analysis of online anonymity and user behaviors: the impact of real name policy. In *IEEE 45th Hawaii International Conference on System Science* (pp. 3041-3050). IEEE. doi: 10.1109/hicss.2012.241.

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70(6).* doi: 10.1103/physreve.70.066111.

Davis, J. L., & Jurgenson, N. (2014). Context collapse: theorizing context collusions and collisions. *Information, Communication & Society, 17(4),* 476-485. doi: 10.1080/1369118x.2014.888458.

Davidson, J., & Martellozzo, E. (2013). Exploring young people's use of social networking sites and digital media in the internet safety context: A comparison of the UK and Bahrain. *Information, Communication & Society, 16(9),* 1456-1476. doi: 10.1080/1369118X.2012.701655

Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer‐Mediated Communication, 15(1),* 83-108. doi: 10.1111/j.1083-6101.2009.01494.x.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6,* 178-190. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences, 106(36),* 15274-15278. doi: 10.1073/pnas.0900282106

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites. *Journal of Computer‐Mediated Communication, 12(4),* 1143-1168. doi: 10.1111/j.1083-6101.2007.00367.x

Facebook (2015) “Annual report” Available at: http://investor.fb.com/secfiling.cfm?filingID=1326801-15-6 Accessed on: 24/04/2015

Facebook[2] (2015) “What names are allowed on Facebook?” Available at: https://www.facebook.com/help/112146705538576 Accessed on 24/04/2015

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior, 25(1),* 153-160. doi:10.1016/j.chb.2008.08.006

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). Social Network Analysis. *Analyzing Social Media Networks with NodeXL*, 31–50. doi:10.1016/b978-0-12-382229-1.00003-5

Hasebrink, U., Görzig, A., Haddon, L., Kalmus, V. and Livingstone, S. (2011). *Patterns of risk and safety online. In-depth analyses from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries*. LSE, London: EU Kids Online. Hayes, A. F. (2015). "The Process Macros for SPSS and SAS" Available at http://www.processmacro.org/index.html Accessed on: 24/04/2015.

Herring, S. C., & Martinson, A. (2004). Assessing gender authenticity in computer-mediated language use evidence from an identity game. *Journal of Language and Social Psychology, 23(4),* 424-446. doi: 10.1177/0261927x04269586

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of suicide research, 14(3), 206-221.* doi: 10.1080/13811118.2010.494133

Hogan, B. (2008). Analysing Social Networks via the Internet. In: N. Fielding, R. Lee and G. Blank (eds) *The Handbook of Online Research Methods*. Thousand Oaks, CA: Sage. doi: 10.4135/9780857020055.n8.

Hogan, B. (2012). Pseudonyms and the rise of the real-name web. *A Companion to New Media Dynamics*, 290-308. doi:10.1002/9781118321607.ch18.

Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2013). Online harassment in context: Trends from three Youth Internet Safety Surveys (2000, 2005, 2010). Psychology of Violence, 3(1), 53. doi: 10.1037/a0030309

Landoll, R. R., La Greca, A. M., Lai, B. S., Chan, S. F., & Herge, W. M. (2015). Cyber victimization by peers: Prospective associations with adolescent social anxiety and depressive symptoms. *Journal of adolescence, 42, 77-86.* doi:10.1016/j.adolescence.2015.04.002

LinkedIn (2015) "Account Security: Your name field" Available at: https://help.linkedin.com/app/safety/answers/detail/a_id/37229/chapter/unacceptable_name Accessed on: 24/04/2015

Litt, E. (2012). Knock, knock. Who's there? The imagined audience. *Journal of Broadcasting & Electronic Media, 56(3),* 330-345. doi: 10.1080/08838151.2012.705195

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). *Pew Research: Teens, social media, and privacy*. Available at: http://www.pewinternet.org/Reports/2013/Teens-Social-Media-And-Privacy/Summary-of-Findings.aspx, Accessed: 21/10/13

Manago, A. M., Taylor, T., & Greenfield, P. M. (2012). Me and my 400 friends: The anatomy of college students' Facebook networks, their communication patterns, and well-being. *Developmental Psychology, 48(2),* 369-380. doi:10.1037/a0026338

Marwick, A. E., & boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13(1),* 114-133. doi: 10.1177/1461444810365313

McCarty, C., Killworth, P.D., Bernard, H.R., Johnsen, E.C., Shelley, G.A. (2001). Comparing two methods for estimating network size. *Human Organization, 60*, 28–39.

New York Times, 2014. "We want privacy, but can't stop sharing" Available at: http://www.nytimes.com/2014/10/05/sunday-review/we-want-privacy-but-cant-stop-sharing.html?_r=0 Accessed: 24/04/2015.

Ofcom (2014) "Adults Media Use and Attitudes Report 2014" Available at: http://stakeholders.ofcom.org.uk/market-data-research/other/research-publications/adults/adults-media-lit-14/ Accessed: 18/03/2015

Pew Research (2014) "Six new facts about Facebook" Available at: http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/ Accessed: 18/03/2015

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40(3)*, 879-891. doi: 10.3758/brm.40.3.879

Rieder, B. (2013). Studying facebook via data extraction: The netvizz application. *Proceedings of the 5th Annual ACM Web Science Conference*, 346-355. doi: 10.1145/2464464.2464475

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., ... & Gleave, E. (2009, June). Analyzing (social media) networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies* (pp. 255-264). ACM. doi: 10.1145/1556460.1556497

Staksrud, E., Ólafsson, K., & Livingstone, S. (2013). Does the use of social networking sites increase children's risk of harm? *Computers in Human Behavior, 29(1),* 40-50. doi: 10.1016/j.chb.2012.05.026

Stefanone, M. A., Lackaff, D., & Rosen, D. (2008, June). We're all stars now: Reality television, Web 2.0, and mediated identities. *In Proceedings of the nineteenth ACM conference on Hypertext and hypermedia (pp. 107-112). ACM.*

Stefanone, M. A., Lackaff, D., & Rosen, D. (2011). Contingencies of self-worth and social-networking-site behavior. *Cyberpsychology, Behavior, and Social Networking, 14(1-2), 41-49*. doi: 10.1089/cyber.2010.0049

Thelwall, M., (2008). Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology, 59(8),* 1321-1330. doi: 10.1002/asi.20835

Underwood, J. D., Kerlin, L., & Farrington-Flint, L. (2011). The lies we tell and what they say about us: Using behavioural characteristics to explain Facebook activity. *Computers in Human Behavior, 27(5),* 1621-1626. doi: 10.1016/j.chb.2011.01.012

Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior, 9(5),* 584-590. doi: 10.1089/cpb.2006.9.584

Vitak, J. (2012). The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media, 56(4),* 451-470. doi: 10.1080/08838151.2012.732140

Washington Post, 2013. "Cyberbullying charges weighed after suicide of Florida girl, 12" Available at: http://www.washingtonpost.com/national/cyberbullying-charges-

weighed-after-suicide-of-florida-girl-12/2013/09/13/65f755fc-1cd6-11e3-a628-7e6dde8f889d_story.html Accessed: 16/07/2015.

Wilcox, K., & Stephen, A. T. (2013). Are close friends the enemy? Online social networks, self-esteem, and self-control. *Journal of Consumer Research, 40(1),* 90-103. doi: 10.1086/668794

YouYou, W., Kosinski, M., & Stilwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *PNAS, 112*, 1036-1040. doi: 10.1073/pnas.1418680112

Table 1

General Sample Characteristics.

	Frequency	%
Gender		
Male	65	36.7
Female	112	63.3
Daily Facebook Engagement		
0-15 minutes	51	28.8
16-30 minutes	45	25.4
31-45 minutes	29	16.4
46-60 minutes	22	12.4
1 hour +	30	16.9
Facebook Privacy		
Don’t Know	9	5.1
Anyone	11	6.2
Friends Only	118	66.7
Friends + Additional Filters	39	22.0

Table 2

Descriptive statistics of self-report and digitally derived measures.

	Mean	Standard Deviation	Minimum	Maximum
Online Vulnerability	2.75	1.09	1.00	5.00
Network Size	399.40	277.25	4.00	1468.00
Network Clustering	.77	.06	.36	1.00
Social ‘Friend’ Types	9.11	2.61	1.00	16.00
Misclassified Profiles	3.18	4.36	.00	27.00
Gender-Hidden Profiles	2.40	3.09	.00	21.00
Pseudonym Profiles	2.49	5.41	.00	57.00
Network Outliers	8.86	11.69	.00	90.00
Age	22.85	9.81	13	77

Table 3

Frequency of social ‘friend’ types reported by the sample (N=177).

Social ‘Friend’ Type	N (%)
Parents	111 (62.7%)
Siblings	137 (77.4%)
Grandparents	44 (24.9%)
Other Family	149 (84.2%)
Best Friend	165 (93.2%)
Friends	175 (98.9%)
Current Classmate	138 (78.0%)
Previous Classmate	152 (85.9%)
Current Teacher/Lecturer	13 (7.3%)
Previous Teacher/Lecturer	54 (30.5%)
Neighbour	50 (28.2%)
Leisure / Interest Group Member	110 (62.1%)
Friend of Friend (FoF)	111 (62.7%)
Casual Acquaintance	109 (61.6%)
Online Only	50 (28.2%)
Celebrities / Public Figures	45 (25.4%)

Table 4

Bivariate Correlations.

	1	2	3	4	5	6	7	8	9	10
1. Online Vulnerability		.383**	-.260**	.370**	.394**	.201**	-.033	.166*	-.104	.143
2. Network Size			-.506**	.430**	.627**	.460**	.271**	.377**	-.139	.165*
3. Network Clustering				-.349**	-.529**	-.441**	-.421**	-.716**	-.370**	-.308**
4. Social 'friend' types					.339**	.326**	.135	.305**	-.006	.268**
5. Misclassified						.516**	.265**	.494**	.081	.213**
6. Pseudonym							.331**	.408**	.077	.035
7. Gender-hidden								.482**	.543**	.135
8. Network outliers									.488**	.339**
9. Age										.241**
10. Gender										

*Note: df =175. *p<.05. **p<.001.*

Table 5

Hierarchical regression analysis.

	DV: Online vulnerability		
	Model 1: Size	Model 2: Diversity	Model 3: Anomalies
Demographics			
Age	-.049 [-.099, -.003]*	-.069 [-.126, -.024]*	-.034 [-.101, .034]
Gender	.074 [-.019, .175]	.028 [-.064, .125]	.006 [-.101, .034]
Network Variables			
Network size	.017 [.010, .024]***	.007 [-.002, .016]	.006 [-.005, .017]
Network clustering		-1.229 [-2.125, -.325]*	-1.112 [-2.190, -.078]*
Social 'friend' types		.132 [.002, .257]*	.146 [.025, .262]*
Network Anomalies			
Misclassified			.069 [.012, .124]*
Gender-hidden			-.055 [1.115, .001]
Pseudonym			-.038 [-.081, .019]
Outliers			.001 [-.042, .039]
Constant	1.487 [1.163, 1.820]***	2.355 [1.172, 3.595]***	2.097 [.875, 3.376]**
	F(3, 176)=12.425***	F(5, 176)=10.733***	F(9, 176)=7.451***
R^2	.177	.239	.287

*Note: *p<.05. **p<.01. ***p<.001. All coefficients are unstandardised.*

Table 6

Analysis of indirect effects (Paths a x b) for Model 1.

							Bootstrapping*	
		Unstandardised Point Estimate	Standardised Estimate	Product of Coefficients		*p*	Bias Corrected 95% CI	
				SE	Z		Lower	Upper
Social types	‘friend’	.004	.078	.002	2.087	.037	.001	.008
Network clustering		.006	.123	.003	2.269	.023	.002	.012

*Note: *Bootstrapping based on 5000 samples.*

Table 7

Analysis of indirect effects (Paths a x b(x d)) for Model 2.

			Bootstrapping*		
	Unstandardised Effect	Standardised Effect	Boot SE	Bias Corrected 95% CI	
Indirect Path				Lower	Upper
1. Size→Friends→Vulnerability	.0039	.078	.0018	.0006	.0079
2.Size→Friends→Cluster→Vulnerability	.0007	.015	.0005	.0001	.0021
3. Size→Friends→Misc.→ Vulnerability	.0000	-.000	.0004	-.0008	.0007
4.Size→Friends→Cluster→Misc.→Vuln.	.0003	.005	.0002	.0001	.0008
5. Size→Cluster→Vulnerability	.0038	.076	.0021	.0001	.0086
6.Size→Cluster→Misc.→ Vulnerability	.0014	.027	.0008	.0002	.0034
7. Size→Misclassified→Vulnerability	.0045	.089	.0021	.0007	.0088

*Note: *Bootstrapping based on 5000 samples. Misc. = Misclassified Profiles. Vuln. = Vulnerability.*

Fig. 1. Facebook network with 269 'friends' and a global clustering of .747. Network Key: red = female; blue = male; black = no gender. Straight lines represent interconnections among ego's Facebook "friends". Curved connections represent multiple, collapsed interconnections among network clusters.

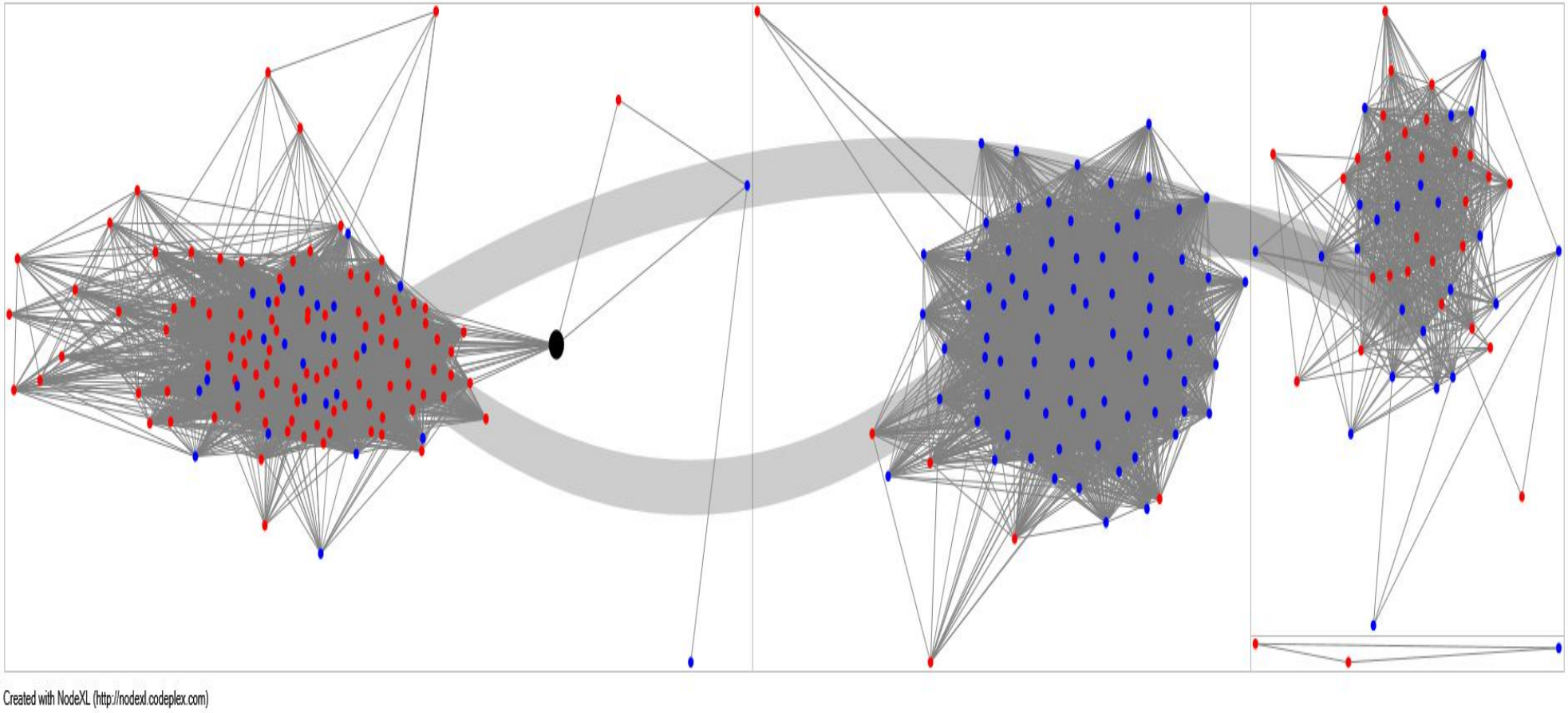

Fig. 2. Facebook network with 235 'friends' and a global clustering of .391. Network Key: red = female; blue = male; black = no gender; solid sphere = typical 'friend'; square = misclassified profile; triangle = pseudonym profile. Straight lines represent interconnections among ego's Facebook "friends". Curved connections represent multiple, collapsed interconnections among network clusters.

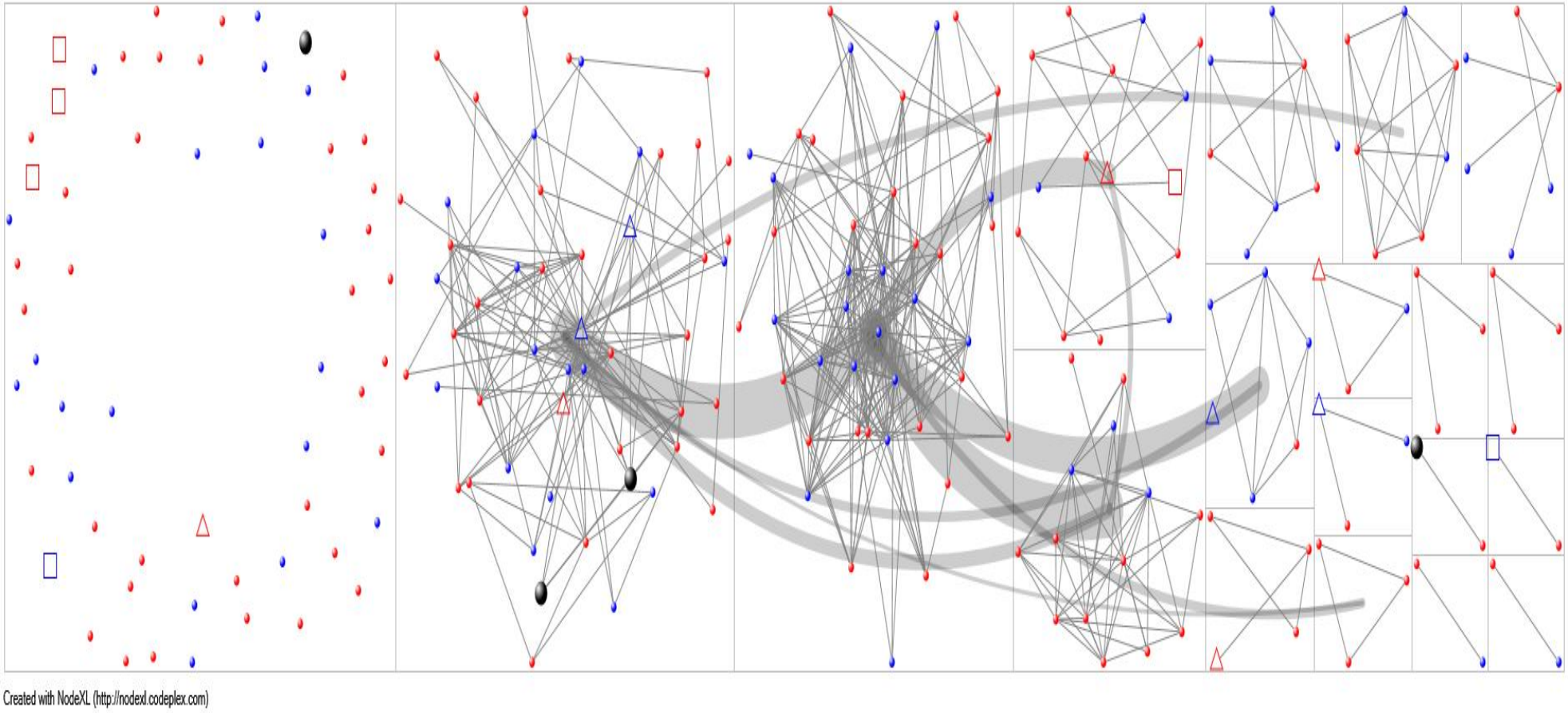

*Fig. 3. Path representation of mediation and effects for Model. Note: *p<.05. **p<.01. ***p<.001. β values represent unstandardised coefficients.*

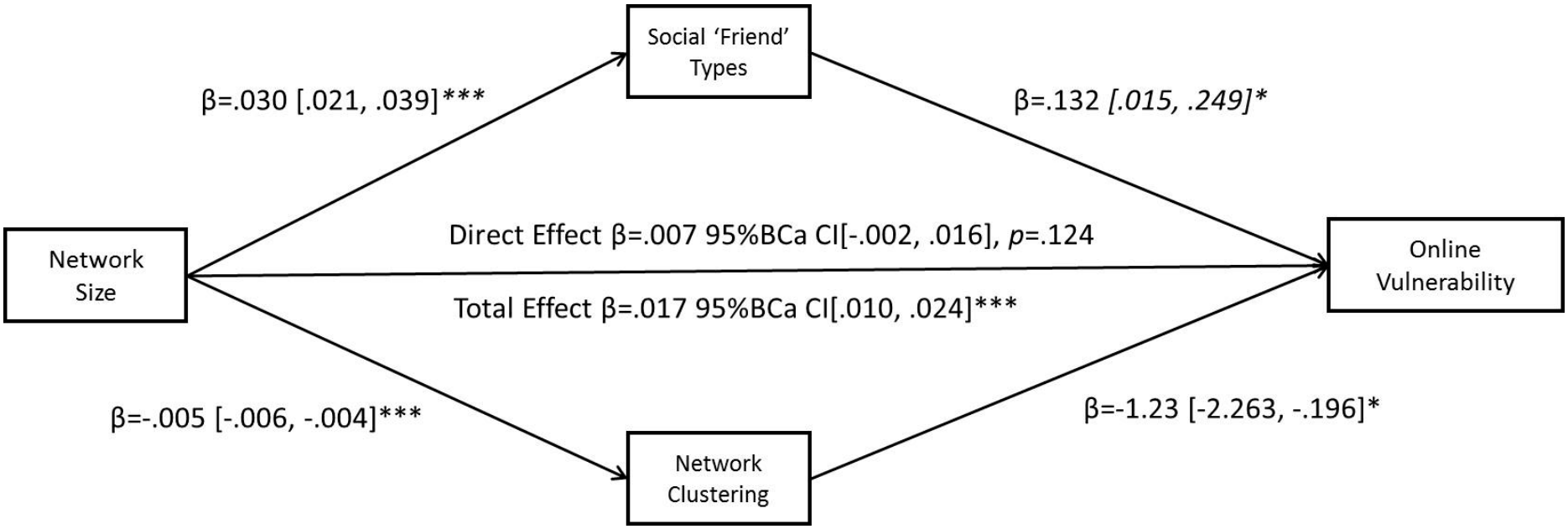

*Fig. 4. Path representation of mediation and effects for Model 2. Note: *p<.05. **p<.01. *** p<.001. β values represent unstandardised coefficients.*