

The Credibility of Digital Identity Information on the Social Web: A User Study

Matthew Rowe
OAK Group
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello Street
S1 4DP Sheffield, United Kingdom
m.rowe@dcs.shef.ac.uk

ABSTRACT

The recent rise in the adoption of Social Web platforms such as MySpace, Facebook and Twitter has provided Web users with rich functionality and feature sets to interact with their peers and construct an online presence. The digital identity which Web users build on the Social Web is being increasingly reused by third party services (product recommendation services, authentication mechanisms, identity management services). The reliance on such digital identity information requires accurate and credible information. This paper presents a detailed user study of the digital identity representations which are constructed on the Social Web. The study explores the extent to which such representations mirror their real-world equivalent and therefore assesses the credibility of such information.

Categories and Subject Descriptors

H.1.1 [MODELS AND PRINCIPLES]: Systems and Information Theory; J.4 [SOCIAL AND BEHAVIORAL SCIENCES]: Sociology

General Terms

Measurement

Keywords

Digital Identity, Information Credibility, Social Web

1. INTRODUCTION

Since its creation the World Wide Web has been changing. It has now evolved from a Web of anonymity - where Web users, unless technologically-savvy, are devoid of presence - into a Web of accountability and visibility. Web users are now able to interact online with Web platforms such as the BBC's news service, write and publish their thoughts

on blogging platforms such as Live Journal, publish videos of their activities on Youtube¹, share photos on Flickr² and interact with their friends on MySpace³. A consistent feature of such platforms and services is their provision of rich functionality and feature sets which allows users to control their persona by handcrafting their own profile as they wish to appear to others. This bespoke persona forms the *digital identity* of the user within the given context or environment.

Digital identity information found on Social Web platforms is being increasingly used by third parties - under the allowance of the profile owner. For instance Web applications such as Dopplr⁴ utilise profile information - specifically social network information - from Social Web platforms to alert Web users of trips being planned by their friends. Social search, such as Google's social search⁵, uses digital identity information of Web users, predominantly their social network, to affect search listings. Digital identity information from the Social Web is now used to aid identity management [17] by supporting the automatic detection of Web pages citing a given person. It is also common for lateral surveillance to be performed using profiles from Social Web platforms to gather an insight into the identity of a given individual through the online persona they maintain. This is practiced by companies vetting prospective employees and socialites who have met someone and wish to look them up online [1].

The digital identity which users construct is beneficial to third parties and an ever increasing number of services and people are beginning to rely on such information. However this begs the question: *is the digital identity which a user constructs on a Social Web platform representative of their real-world identity?* In answering this question one must consider the credibility of digital identity information found on the Social Web. Given that Web users are responsible for their own Web appearance and are able to shape their online persona and how they are viewed by others, an investigation is required which assesses the extent to which the Web users' digital identities mirrors their real-world identities.

Our contribution within this paper is a detailed user study of the credibility of digital identity information found on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WICOW'10, April 27, 2010, Raleigh, North Carolina, USA.
Copyright 2010 ACM 978-1-60558-940-4/10/04 ...\$10.00.

¹<http://www.youtube.com>

²<http://www.flickr.com>

³<http://www.myspace.com>

⁴<http://www.dopplr.com>

⁵<http://googleblog.blogspot.com/2009/10/introducing-google-social-search-i.html>

the Social Web. To motivate our work we investigate the following hypothesis:

Social Web users construct digital identity representations which mirror their real-world identities.

The presented study compares the social networks which users of Facebook⁶ maintain offline with their digital equivalent. We use adapted information retrieval measures to measure the *coverage* and *relevance* of the digital social network with respect to its real-world equivalent. Our results empirically demonstrate a significant overlap between the two networks and therefore indicates that to a large extent the digital identities which Web users construct through the feature sets of Social Web platforms mirror real-world identity.

We have structured the paper as follows: section 2 discusses the notion of digital identity and how Web users are motivated to shape their identity on Social Web platforms. Section 3 discusses related work which investigates the comparison of digital and real-world identity information, placing particular emphasis on work performed within the sociology community. Section 4 presents our user study in which we explain the experimental setup, the participants used, the measures used to quantify the comparison of digital and real-world identity information and a discussion of the findings from the study. Section 5 concludes the paper with the lessons learnt from this work with respect to the credibility of digital identity information found on the Social Web.

2. DIGITAL IDENTITY

The established dictionary definition of identity is “*the fact of being who or what a person or thing is*” and “*the characteristics determining this*”⁷. This definition indicates that the identity of a person is comprised of a set of attributes or properties which make the person unique. Digital identity follows this same notion, however within the context of the Social Web the identity of users is bespoke and can be altered by the individual which it describes. Such alterations are possible through functionalities and feature sets on Social Web platforms such as profile pages. On such pages users are able to create an identity profile consisting of their biographical information (name, picture, date of birth, location) and build a social network by adding friends.

Profile pages have altered the visibility of many Web users from being anonymous into an accountable presence. Work in [10] described a next generation Internet involving Web users not merely as entities utilising services, but as citizens who are given civil rights, a digital profile, and aid with establishing trust between other Web citizens through vouching. One can see how the social nature of the Web of today has fulfilled some of these requirements: by giving regular Web users a visible online presence and trusting those users to create content Web users are no longer consumers but are now citizens. This increased visibility of Web users has produced the need for users to control how they are perceived online.

Information relating to digital identity can be divided into three tiers as described in [21] and shown in Figure 1. The first tier is called *My Identity*, which contains persistent identity information such as a person’s name, date

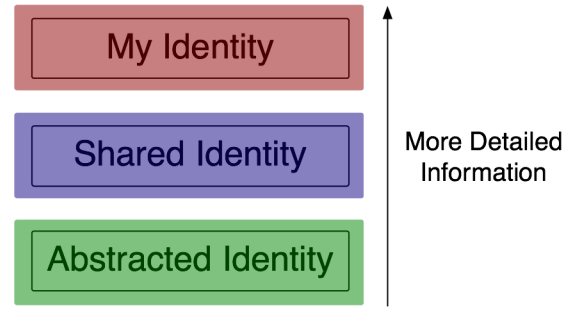


Figure 1: The tiers of Digital Identity

of birth and genealogical relations, essentially information that is constant and is unlikely to change. The second tier is called *Shared Identity*, which contains attributes assigned to an individual by other people such as the social network of a person. *Shared Identity* contains information which is susceptible to change as a person makes friends with different people and loses contact with others. The third tier is called *Abstracted Identity* and contains identity information derived from groupings and demographics. For example, identifying a person by a community of practice which they are involved in. Identity information within this tier is very likely to alter as a person’s interests evolve over time. As Figure 1 demonstrates as the tiers move up the information describing the digital identity of an individual becomes increasingly detailed and therefore deterministic (in the sense of uniquely identifying the person).

Recent work discussing the motivational aspects for building a digital identity [16] states that self-promotion plays a key role by allowing users to create a profile which represents their real-world self as they wish to be perceived. Studying the behavioural characteristics of Social Web users, work by [12] found that users would use Social Web platforms to search for existing people they know and interact with those individuals rather than making new friends. According to [5] the social networks which users construct on Social Web platforms consist of both *strong-tied* and *weak-tied* relationships. The former are relationships which are driven by frequent interactions (i.e. close family and friends) whereas the latter are the opposite: infrequent interactions and are often relationships with acquaintances.

The exposure of social network information by Web users, according to [5], allows third parties to validate the identity of an individual: in essence identifying a person by the company they keep. Such intuition is also used extensively within the field of identity disambiguation - the practice of discovering Web pages which refer to a given person, the process of which is compounded by the ambiguity of person names. For instance work by [9, 11, 13] employ the intuition that a person will co-occur in a Web page with those people he/she knows. Such approaches indicate that the comparison of identity representations can be performed through the analysis of social network information - given that such information belongs to the *shared identity* tier of digital identity information.

3. RELATED WORK

Several pieces of work have assessed the relation between digital and real-world identities. For instance work by [2]

⁶<http://www.facebook.com>

⁷http://www.askoxford.com/concise_oed/identity

details a user study in which a comparison of online and offline social networks are made to derive similarities and differences in relationship qualities. The results from the study show that offline social networks are largely strong-tied and consist of frequent interactions, whereas online relationships require less involvement and the majority are weak-tied: similar to findings in [5]. A user study described in [3] explores the use of Social Web platforms by adolescents, particularly with respect to building an online identity. The findings from the study indicate that the participants used Social Web platforms to maintain existing relationships and that users built digital personas to reflect their real-world identities. An extensive user survey is presented in [15] which analyses the usage of Social Web platforms in relation to social networking. Of the participants queried 69% used such Social Web platforms to talk to friends and family that they are close to offline, 65% talk to friends and family that they rarely speak to offline and 17% used Social Web platforms to talk to people they do not know offline.

Although the above work is useful by providing an insight into the dynamics of digital identity, and in particular the social networks which users maintain, on the Social Web it does not explicitly quantify the similarity between digital and real-world identity information. Work by [19] moves towards this area goal by performing social network analysis over a college social network gathered from Facebook. Analyses made over this data show correlations between online social networks and real-world social networks - i.e. communities of practice within the college environment reflect online social networks and cliques. However the findings are not deterministic such that there is no definitive proof that the online social networks are indeed valid, this being especially pertinent given the propensity of online social networks to be largely composed of weak-tied relationships [5].

Online and offline social networks are compared in [18] through a detailed user study. Participants of the study listed their top ten face-to-face friends - therefore resembling their real-world social network - and their top 10 friends on instant messaging programs and Social Web platforms - thereby giving their digital social network. The authors then compared the overlap between these two networks. The results from the study found that 49% of participants' real-world social network was also present online and that 22% of the participants (16 people) had a online social network which contained all their listed real-world relationships.

4. COMPARING DIGITAL AND REAL-WORLD IDENTITIES VIA SOCIAL NETWORK SIMILARITY

The declarative qualities of social network information form an important and defining aspect of the digital identity which users construct on Social Web platforms. The use of such information in the context of identity disambiguation and the user studies described within the previous section suggest that digital identity and real-world identity information can be compared using the similarity between social networks: online and offline. Such a comparison would then quantify the extent to which digital identity information is representative of real-world identity information and furthermore, provide an important insight into the credibility of digital identity information on the Social Web.

To provide the analysis of real and digital identity information a Social Web user study was conducted using the Social Web platform Facebook. Facebook has now become the most widely used Social Web platform in the UK - the country in which the user study was conducted - with 22 million users⁸. The remainder of this section describes the experiment used for the study in detail, explaining the selection of participants, the experimental procedure, the empirical measures used for quantitative analysis and the results.

4.1 Experimental Setup

The user study consisted of an experiment designed to provide an insight into the repetition of offline social networks in an online environment. The behaviour of Social Web users was also analysed to investigate which members of their online social network participants interacted with the most. Participants for the user study were gathered by emailing all students and staff at the University of Sheffield requesting their participation in the experiment. It was specified that each participant must have an account with the Social Web site Facebook. In total the user study consisted of 50 participants - 25 male and 25 female - covering a wide age range (18 - 45). The experiment was conducted in November 2008 and was comprised of 3 separate steps as follows:

4.1.1 Stage 1. Real-World Social Network Creation

The first step involved each participant describing his/her social network in the real-world. They were given a Web page form with 20 rows, one for each person in the participant's social network. Each row was completed with the name of a person, their age and the type of relationship (family, friend, colleague). More rows could be added should the participant wish to add more than 20 people. The intention of this step is for the participant to list his/her social network containing the strongest and most important (strong-tied) relationships [5], given that these will be the first that come to mind.

4.1.2 Stage 2. Digital Social Network Extraction

The social network of each participant was extracted from Facebook using the Social Circular Facebook application⁹, specifically developed for the study. Image and conversation data was analysed for behavioural trends: images were extracted that each participant featured in. For each participant the images were analysed to find which other people appeared within the images. In order to analyse the conversation data, all messages each participant shared were extracted. In a similar manner to the image analysis, the messages were analysed to find which people had sent and received them.

4.1.3 Stage 3. Comparing the Real World and Digital Social Networks

The final stage of the procedure involved the comparison of the two social networks. Each participant compared his/her real-world social network from stage 1 with the digital social network from stage 2. If a person appeared in both the real-world and digital networks then the participant flagged that person. In comparing the networks, the

⁸<http://www.clickymedia.co.uk/2009/10/uk-facebook-user-statistics-october-2009/>

⁹<http://apps.facebook.com/socialcircular>

overlap in people present in both networks is derived and therefore gives a quantifiable measure of the replication of offline social networks in an online environment.

4.2 Evaluation Measures

To measure the extent of the replication of real-world social networks in an online environment, the information retrieval metrics precision and recall as described by [20] were used. Information retrieval metrics are designed for measuring the accuracy of information discovery systems and the classic definitions feature two sets of documents: A denotes the set of *relevant documents* and B denotes the set of *retrieved documents*. Therefore precision and recall are defined as follows:

$$precision = \frac{|A \cap B|}{|B|} \quad (1)$$

$$recall = \frac{|A \cap B|}{|A|} \quad (2)$$

Such information retrieval metrics are well suited to the experimental setup used for this study. Imagine that the set of relevant documents, A , actually refers to the set of *relevant people*. When each participant defines his/her real-world social network, this network characterises the most relevant people to that person. [20] defines relevance as being assumed to have a broader meaning of "aboutness" and "appropriateness" which given the context of the study suits the relevance of people to the participant. In a similar manner, one can consider the set of retrieved documents, B , to actually denote the set of *retrieved people*, which in the context of the study is the digital social network. Based on these adaptations of the document sets, the metrics are redefined as relevance and coverage as follows:

4.2.1 Relevance

The definition of precision from [20] describes a measure which returns "...the proportion of retrieved information that is relevant...". Given that in this study relevant information is the real-world social network and retrieved information is the digital social network, precision can be redefined as a measure of the *relevance* of a participant's digital social network as follows:

$$relevance = \frac{|real \cap digital|}{|digital|} \quad (3)$$

This metric of relevance therefore measures the ratio of strong-tied to weak-tied relationships within the digital social network, or rather, the proportion of the digital social network that is denoted by strong-tied relationships and therefore relevant. For instance, it is common for Social Web users to maintain a social network containing many weak-tied relationships [5]. As each study participant lists his/her real-world social network it is likely that they will list strong-tied relationships, given that these are the first that come to mind. Using the above metric derives a measure of relevance of the digital social network with respect to the real-world social network.

4.2.2 Coverage

The definition of recall from [20] defines a measure which returns "...the proportion of relevant material that is actually

retrieved...". Given that relevant material is the real-world social network and retrieved material is the digital social network, recall is redefined as a measure of *coverage* as follows:

$$coverage = \frac{|real \cap digital|}{|real|} \quad (4)$$

Coverage provides a measure of the extent of the replication of each participant's real-world social network within a virtual environment. The intersection is derived between the set of people from the real-world social network and the set of people from the digital social network. The size of this intersection is then normalised by the size of the real-world social network. The derived value returns a quantifiable measure of the proportion of the real-world social network that is now replicated online, and thus to what extent the real-world social network is covered by the digital social network.

4.3 Results

4.3.1 Relevance and Coverage

The results from the user study (presented in Figure 2) show that, in general, the majority of each participant's real-world social network is replicated in an online environment. Coverage provides a measurement of the extent of this replication, where the results show a range of coverage from 0.5 to 1 - where only 1 person out of the 50 participants achieved 1 - with an average coverage measure of 0.77. To place this in context, on average 77% of a given Social Web user's real-world relationships are repeated in an online environment. This figure indicates that digital identity constructed on a Social Web platform is now, to a large extent, representative of a Web user's identity in the real world.

Such findings can be directly compared with those from work by [18] which found that on average 49% of a Social Web user's offline social network was replicated online. The difference between the two values could be attributed to the environment in which the user study participants were involved: in [18] the Urban University of Los Angeles was used as the participant source whereas in the case of this study the University of Sheffield, therefore the differences in culture and attitude towards Social Web platforms may have influenced the results. Another difference is the time-liness of when the studies were conducted, the growth in the number of Social Web platforms has lead to their establishment as a means of communication and a norm within society. Therefore the recent results presented within this paper suggest the tendency of users to utilise Social Web platforms to mimic their real-world identities.

The results from the study also produced an average relevance measure of 0.23. To contextual this result, on average 23% of a given Web user's online social network consists of important, strong-tied relationships and the remaining 77% are weak-tied relationships. One can imagine that online social networks, although intrinsically social in nature, do not contain a large number of important relationships. [7] argues that relationships found on Social Web sites form part of much vaster social networks, and are mainly classed as weak-ties; where the two people are acquainted and nothing more.

As the results in Figure 2 demonstrate coverage tends towards 0.9 as the online social network increases in size. Al-

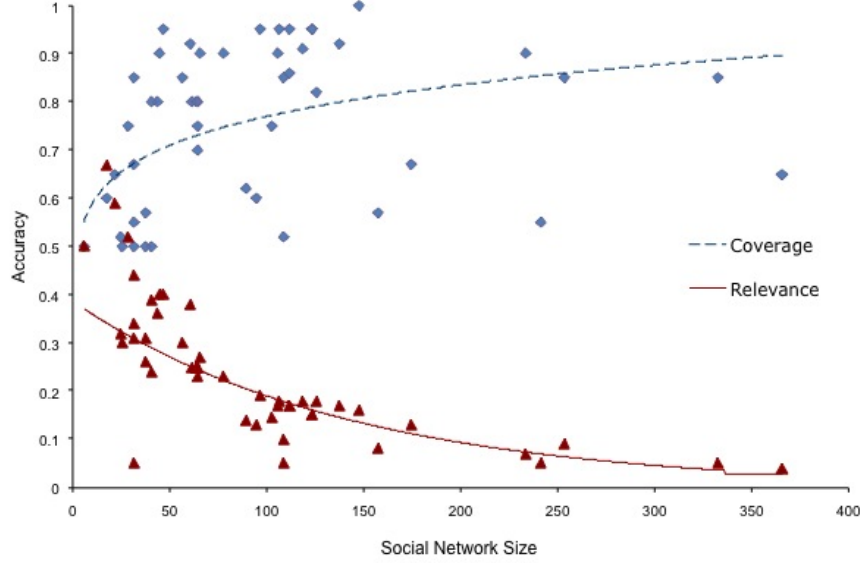


Figure 2: Relevance and Coverage measures from social network comparisons

though the adoption of Social Web platforms by Web users has become a mainstream phenomena, there are still users who resist following their friends and peers in signing up to such services. We observed a nonlinear relationship between the size of the social network and both relevance and coverage. The results also show a logarithmic regression [4] model best fits the trend of the data: a nonlinear relationship exists between the size of the social network and the level of coverage. The larger the digital social network becomes, the more relationships it contains, strong-tied and weak-tied, therefore the likelihood of the real-world social network overlapping with the digital social network increases. Conversely, when the social network grows in size relevance is reduced: as the number of weak-tied relationships within the network gradually increases, the ratio of strong-tied relationships to weak-tied relationships grows.

4.3.2 Demographics of Coverage

Deeper exploration of the collected data also provides an analysis of the portion of the real world social network which is replicated online. We queried this portion of the social network for the relationships types which it consisted of. Recall from the description of the experiment that when compiling their real-world social networks, participants were given a form in which they selected the type of relationship which they had with their social network member - choosing from either *coworker*, *friend* or *family*. Figure 3 shows the breakdown of the distribution of these three relationship types. The majority of real-world relationships which are replicated online - and are therefore strong-tied - were with friends (62%). Family and friends both made up much smaller portions of the covered social network - found to be 24% and 14% respectively.

Further analysis of the data found that for 68% of relationships from the real-world social networks which were with coworkers were replicated online. For friends 89% of strong-tied relationships were maintained in a digital social network

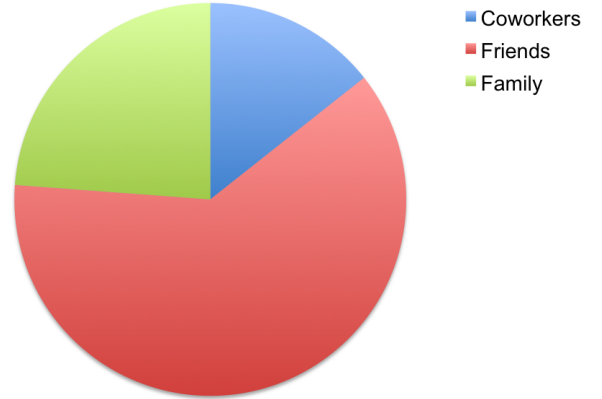


Figure 3: Distribution of relationship within the overlap between real world and digital social networks

and 73% of family relationships listed in the real-world social network were replicated online. Such findings suggest that the demographic of strong-tied relationships within a digital environment are largely driven by interactions with friends rather than coworkers or family members.

Using the collected data from the experiments we also analysed the age demographic within the portion of the real-world social networks which were covered online. Figure 4 shows the breakdown of these ranges. Given the participants within the study were largely students who used Facebook regularly on campus, it is unsurprising that the largest age group within the covered portion of the network consisted of the lowest age band tested (< 21). As the age increases the portion of the real-world social network replicated online decreases.

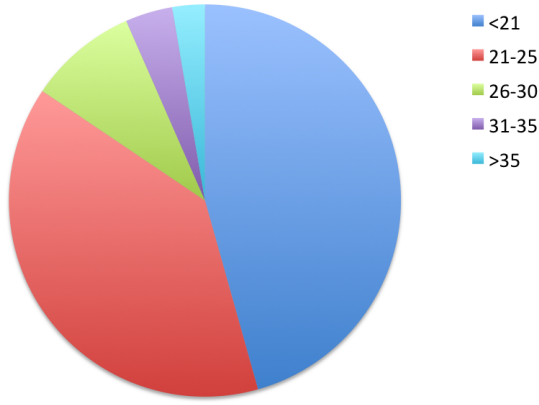


Figure 4: Distribution of ages within the overlap between real world and digital social networks

4.3.3 Behaviour

The behaviour of each participant within their social network was analysed by collecting image and message data. Figures 5(a) and 5(b) present the results of the analysis. Images featuring each participant were analysed to see which other people appear in those images, therefore deriving a cumulative count for each member of the participant’s social network. A similar approach was used for analysing the messages sent and received by each participant to count how many times each social network member had shared a message with the participant.

The results from this analysis demonstrate a trend in the behaviour of the participants to share a low number of messages and images with a large number of people. This interaction could be contextualised as a Web user sharing a rare conversation with an acquaintance. The results also show how each participant shares a high number of images and messages with very few people. Figures 5(a) and 5(b) demonstrates how the data forms a power law [14] curve where the head of each graph contains a high frequency of people with whom very few images and messages are shared, whereas the tail contains a low frequency of people with whom many images and messages are shared. We refer to this structure as the *Social Longtail*.

The social longtail represents those people within a person’s online social network with whom the person has a relationship incorporating frequent interactions. Those relationships are maintained by the need to communicate in an online environment by sharing images and sending and receiving messages. When analysing the longtail of each graph (>15 for images and messages) on average 92% of the people who feature in the longtail are members of the real-world social network. This indicates that the social longtail is representative of a given Web user’s real-world social network and contains strong-tied relationships. The real-world social network contains the people with whom the participants interact the most and have the strongest relationships - bolstered by these frequent interactions - the large extent to which these people appear within the social longtail indicates that the interactions driving offline relationships are repeated online. This indicates that the participants use Social Web platforms as an extension of their real-world identities and maintain a similar persona and social network.

Such findings present a detailed insight into the digital identity which Web users construct on the Social Web in relation to their real-world equivalent. The large extent to which real-world relationships are mirrored in online social networks is consistent with existing work exploring the similarity between online and offline social networks [2, 3, 18, 19] and provides a confirmation of theoretical discussions made in [6, 8] that social networks created and maintained on Social Web platforms are merely extensions of their offline equivalents.

5. CONCLUSIONS

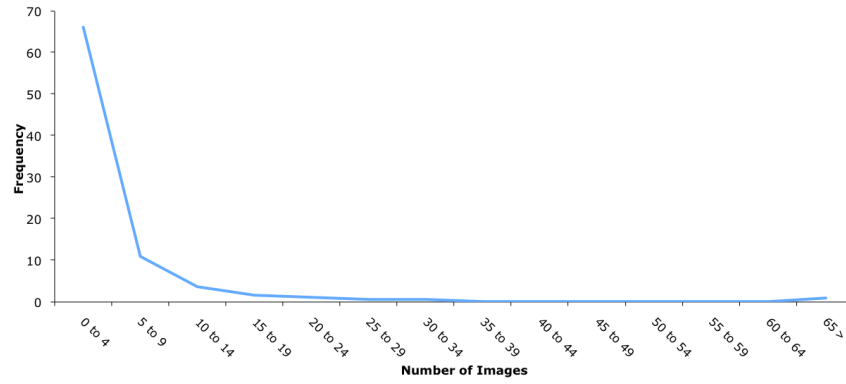
The credibility of digital identity information on the Social Web is important to many Web-based applications that consume such information. The increased reliance on digital identity information could have detrimental effects if the identity representation which Web users build on Social Web platforms are not representative of their real-world identities - particularly within the domain of identity disambiguation. This paper has presented an extensive user study comparing digital and real-world identity information via social network similarity. The results from the study demonstrate that online social networks, which form an intrinsic part of the digital identity of Web users, replicate a large portion Web users’ offline social networks. In short we believe that the presented empirical evidence proves our earlier hypothesis:

Social Web users construct digital identity representations which mirror their real-world identities.

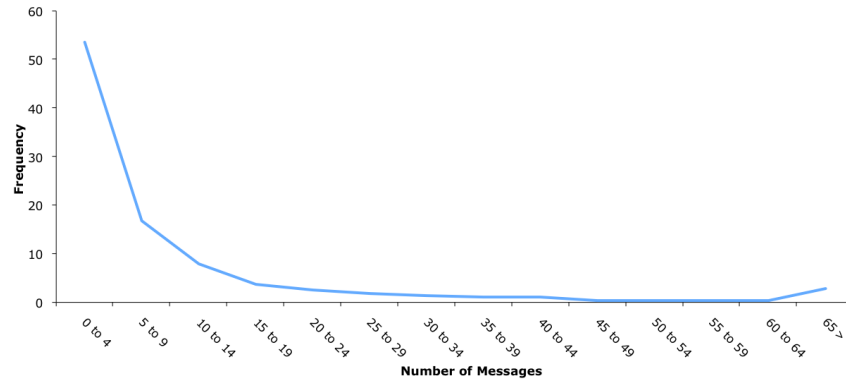
The work presented within this paper has provided an insight into the digital identity which users construct on the Social Web. The study and comparison of online and offline social networks suggests that digital identity information found on the Social Web is credible and reflects the real-world identity of Web users. Of course our study only considers a single Social Web platform. Future work will address this current limitation by extending the scope of analysis to cover additional Social Web platforms and therefore provide additional results from which credibility can be gauged.

Our analysis of the behavioural trends of the participants indicates that the relationships and frequency of interactions which are present offline are repeated in an online space. The *social longtail* of Social Web platform usage contains the majority of strong-tied real-world relationships which are equally as pertinent in an online space. This phenomena suggests that an assessment of the credibility of digital identity information could be facilitated through an analysis of those relationships with respect to the real-world social network, thereby ignoring the remainder of the digital social network of which the majority is weak-tied.

One of the novel contributions within this paper has been the presentation of two distinct metrics: *coverage* and *relevance*, which were used to measure the similarity between social networks. Although [18] presents work which is closely related to our user study, we believe that the use of only a measure of overlap does not provide additional insights into comparisons between real-world and digital social networks. Instead our provision of the described metrics, coupled with our experimental setup, will help future work within the area of digital identity credibility and allow the assessment of the credibility of social network information.



(a) Images



(b) Messages

Figure 5: Analysis of information sharing within digital social networks

6. ACKNOWLEDGEMENTS

The author would like to thank the participants from the University of Sheffield who were involved in the user study. Their time and effort has been invaluable in collecting the results presented within this paper and providing the assessment of the credibility of digital identity information on the Social Web. The research leading to these results has received funding from the EU project WeKnowIt10 (ICT-215453).

7. REFERENCES

- [1] M. Andrejevic. The discipline of watching: Detection, risk, and lateral surveillance. *Critical Studies in Media Communication*, 23(5):392–407, 2006.
- [2] D. K.-S. Chan and G. H.-L. Cheng. A Comparison of Offline and Online Friendship Qualities at Different Stages of Relationship Development. *Journal of Social and Personal Relationships*, 21(3):305–320, 2004.
- [3] B. Clarke. Friends forever: How young adolescents use social-networking sites. *IEEE Intelligent Systems*, 24:22–26, 2009.
- [4] R. Darlington. *Regression and Linear Models*. McGraw-Hill, 1990.
- [5] J. Donath and D. Boyd. Public displays of connection. *BT Technology Journal*, 22(4):71–82, October 2004.
- [6] N. B. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook friends: Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, 12:1143–1168, 2007.
- [7] R. Gross, A. Acquisti, and H. J. Heinz, III. Information revelation and privacy in online social networks. In *WPES ’05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, New York, NY, USA, 2005. ACM.
- [8] J. Hart, C. Ridley, F. Taher, C. Sas, and A. Dix. Exploring the facebook experience: a new approach to usability. In *NordiCHI ’08: Proceedings of the 5th Nordic conference on Human-computer interaction*, pages 471–474, New York, NY, USA, 2008. ACM.
- [9] L. Jiang, J. Wang, N. An, S. Wang, J. Zhan, and L. Li. Two birds with one stone: A graph-based framework for disambiguating and tagging people names in web search. In *18th International World Wide Web Conference (WWW2009)*, April 2009.
- [10] K. Jordan, J. Hauser, and S. Foster. The augmented

- social network: Building identity and trust into the next-generation internet. *First Monday*, 2006.
- [11] D. V. Kalashnikov, Z. Chen, S. Mehrotra, and R. Nuray-Turan. Web people search via connection analysis. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1550–1565, 2008.
 - [12] C. Lampe, N. Ellison, and C. Steinfield. A face(book) in the crowd: social searching vs. social browsing. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 167–170, New York, NY, USA, 2006. ACM.
 - [13] B. Malin. Unsupervised name disambiguation via social network similarity. In *Workshop on Link Analysis, Counterterrorism, and Security*, 2005.
 - [14] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323, 2005.
 - [15] Ofcom. Social networking: A quantitative and qualitative research report into attitudes, behaviours and use. Technical report, Ofcom, April 2008 2008.
 - [16] B. Ploderer, S. Howard, and P. Thomas. Being online, living offline: the influence of social ties over the appropriation of social network sites. In *CSCW '08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 333–342, New York, NY, USA, 2008. ACM.
 - [17] M. Rowe. Applying semantic social graphs to disambiguate identity references. In *Proceedings of the 5th European Semantic Web Conference*, LNCS, pages 461–475. Springer-Verlag, 2009.
 - [18] K. Subrahmanyam, S. Reich, N. Waechter, and G. Espinoza. Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, 29(6):420–433, November 2008.
 - [19] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter. Community structure in online collegiate social networks. Technical Report arXiv:0809.0690, American Physical Society, Sep 2008.
 - [20] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 2nd edition, 1979.
 - [21] P. Windley. *Digital Identity*. O'Reilly Media, Inc., August 2005.