UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

The Impact of Consolidating Web Based Social Networks on Trust Metrics and Expert Recomendation Systems

by

Muhammad Imran

Thesis for the degree of Doctor of Philosophy

September 2015

ABSTRACT

Individuals are typically members of a variety of web-based social networks (both explicit and implied), but existing trust inference mechanisms typically draw on only a single network to calculate trust between any two individuals. This reduces both the likelihood that a trust value can be calculated (as both people have to be members of the same network), and the quality of any trust inference that can be drawn (as it will be based on only a single network, typically representing a single type of relationship). To make trust calculations on MUltiple DIstributed (MuDi) social networks, those networks must first be consolidated into a single network.

Two challenges that arise when consolidating MuDi networks are their heterogeneity, due to different name representation techniques used for participants, and the variability of trust information, due to the different trust evaluation criteria, across the different candidate networks. Semantic technologies are vital to deal with the heterogeneity issues as they permit data to be linked from multiple resources and help them to be modeled in a uniform representation using ontologies. The inconsistency of multiple trust values from different networks is handled using data fusion techniques, as simpler aggregation techniques of summation and weighted averages tend to distort trust data.

To test the proposed semantic framework, two set of experiments were run. Simulation experiments generated pairs of networks with varying percentages of Participant Overlap (PO) and Tie Overlap (TO), with trust values added to the links between participants in the networks. It analysed different data fusion techniques aiming to identify which best preserved the integrity of trust from each individual network with varying values of PO and TO. A real world experiment used the findings of the simulation experiment on the best trust aggregation techniques and applied the framework to real trust data between participants that was extracted from a pair of professional social networks. The trust values generated from consolidated MuDi networks were then compared with the real life trust between users, collected using a survey, with the aim of analysing whether aggregated trust is closer to real life trust than using each of the individual networks.

Analysis of the simulation experiment showed that the Weighted Ordered Weighted Averaging (WOWA) data fusion technique better aggregated trust data and, unlike the other techniques, preserved the integrity of trust from each individual network for varying PO and TO ($p \leq 0.05$). The real world experiment partially proved the hypothesis of generating better trust values from consolidated MuDi networks and showed improved results for participants who are part of both networks ($p \leq 0.05$), while disproving the claim for those in the cross-region (with one user present in both networks and the other in a single network) and single-network users ($p > 0.05$).

Contents

List of Figures

List of Tables

List of Algorithms

Declaration of Authorship

I, Muhammad Imran , declare that the thesis entitled *The Impact of Consolidating Web Based Social Networks on Trust Metrics and Expert Recomendation Systems* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- simulation portion of this work has been published as: a single conference paper in ASE Human Journal Imran et al. (2012)

Signed:..

Date:..

Acknowledgements

Firstly, I would like to express my sincere gratitude and appreciation towards my supervisors Dr. David Millard and Dr. Thanassis Tiropanis for their continuous support throughout the course of my PhD. It was their consistent guidance and motivation that has helped me to achieve this milestone.

I would also like to extend my special thanks to the Rector and management of the COMSATS Institute of Information Technology for providing me an opportunity to study at the world-class University. Without their financial support it would have been impossible for me to undertake this study.

My special thanks also goes to all those lab mates, academia members who have participated in the trust survey conducted as part of my research. It was their valuable data that has empowered me to defend my work at the end.

Last but not least, Thanks to my parents, siblings and other family members for providing me moral support at every stage of the degree, and my life in general.

Nomenclature

$N1$	Original network 1
$N2$	Original network 2
$MuDi$	Multiple distributed networks
$CN1$	Sub-network representing N1 in consolidated MuDi networks
$CN2$	Sub-network representing N2 in consolidated MuDi networks
N	Number of participants in N1
D	Density of N1
C	Average clustering coefficient of N1
L	Average length of shortest paths in N1
N_p	A pair of participants
PO	Participant overlap
TO	Tie overlap
$N_p^{overlapping}$	An overlapping pair of participants
$N_p^{cross-region}$	A cross-region pair of participants
$N_p^{single-network}$	A single-network pair of participants
URI	Uniform resource identifier
TS	Average strength of trust ties
TL	Average length of trust paths
$MuDiTCF$	Multiple distributed trust consolidation framework
RDF	Resource description framework
$RDFS$	Resource description framework schema
$FOAF$	Friend of a friend network ontology
OWL	Ontology web Language
$SPARQL$	Query language for mining data from the rdf triplestore
T_{N_p}	Set of n trust values available from n social networks
T^{N_p}	Aggregated trust value generated from set of values T_{N_p}
S	Summation
WA	Weighted average
$WOWA$	Weighted ordered weighted averaging
$IOWA$	Induced ordered weighted averaging
p_{WA}	Weight vector representing trust source reliability in WA
w_{WOWA}	Weight vector representing information importance in WOWA

p_{WOWA}	Weight vector representing source reliability in WOWA
w_{IOWA}	Weight vector representing information importance in IOWA
p_{IOWA}	Weight vector representing source reliability in WOWA
$ePrints$	A publication archive managed at University of Southampton
$WAIS$	Web and internet science research group at University of Southampton
$TAEL_{ePrints}$	Trust-aware expert list from ePrints
$TAEL_{wais}$	Trust-aware expert list from WAIS
$TAEL_{mudi}$	Trust-aware expert list from multiple distributed social networks

Chapter 1

Introduction

1.1 Motivation

Online social networks (Garton et al., 1997; Boyd and Ellison, 2007) are a modern alternative to offline social networks in which, unlike face-to-face encounters, people interact with each other via the web, potentially anonymously. In these networks, activities and interactions that were once subject to physical presence are now possible electronically, from different locations. Results from the Pew Research survey[1] conducted in January 2014 reveal that 74 per cent of the adult web users use online social networks for interaction (see Figure 1.2(a)). Due to increasing use, issues arise in online social networks that do not arise in real-life social networks (Shariff and Zhang, 2014). For example, people can fake identities (Bilge et al., 2009)), misuse personal information by breaking into the profiles of other users or hack into the online accounts of fellow users to spread false information (Doerr et al., 2012). These are virtually impossible in offline social networks due to the absence of the web, the layer that connects users virtually.

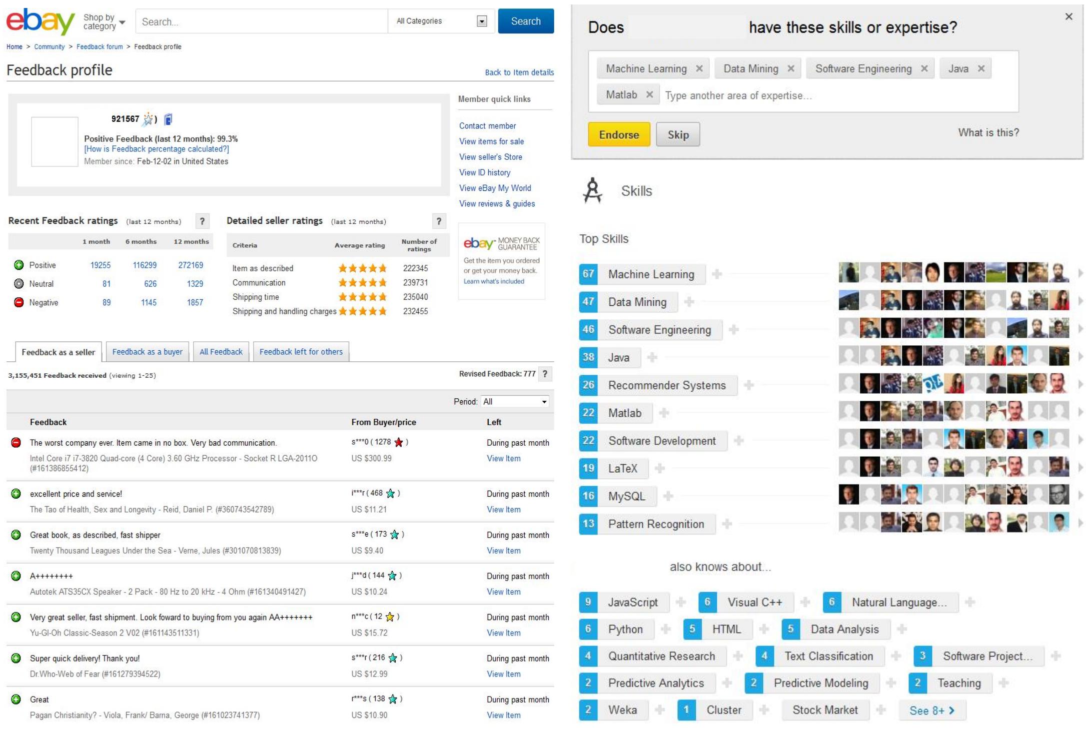
To overcome these problems, researchers have developed trust mechanisms that work with individual social networks across the web. These are based on social and psychological theories of trust derived from human behaviour in the real world (Lewis and Weigert, 1985). A definition of trust borrowed from Olmedilla et al. (2006) describes the perspective of trust used in this work:

> *Trust of party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period (within a specified context in relation to service X).*

Trust models in online social networks analyse the personal information and the interaction history of users to assess their reputation in the respective network. Any novice to the network can use the information as a guide to whether or not to trust another user.

[1] http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

There are many working examples on the web that use the phenomenon of trust, for instance the eBay[2] mechanism to evaluate the reputation of sellers (see Figure 1.1(a)). Buyers rate sellers on the basis of service delivery. Potential buyers then use this information as criteria to decide whether or not to buy from that particular seller. This reputation-based mechanism impacts on the future business of sellers so compels them to satisfy the expectations of users. Another example of such a model is the expertise recommendation in the LinkedIn network (shown in Figure 1.1(b)). Professionals in the network endorse other users on the basis of their areas of expertise. As a result, each professional attains a trust score corresponding to each area, representing the number of professionals who have declared that person as an expert in that area. Any stranger can analyse trust scores to decide whether or not to trust a professional in the corresponding area of expertise.



(a) Ebay reputation of a sample seller. (b) Linked data expert recommendation of a sample professional.

Figure 1.1: Two already developed reputation and expert recommendation systems in ebay and LinkedIn networks.

Importance of trust in online social networks increases significantly when individuals belong to MUltiple DIstributed (MuDi) social networks, with different virtual identities in each network. The term 'MuDi' is used in this thesis to represent multiple social networks on the web, because they are largely owned by different enterprises with information about users is available at distributed locations. These networks provide a

[2]http://www.ebay.com

variety of services not otherwise available through any single network. This increases the likelihood of malicious behaviour, because now users have multiple identities for initiating interaction. A survey[3] (shown in Figure 1.2(b)) has revealed that 42 per cent of social network users are part of MuDi social networks (detailed percentage of overlap between different networks is given in Appendix A).

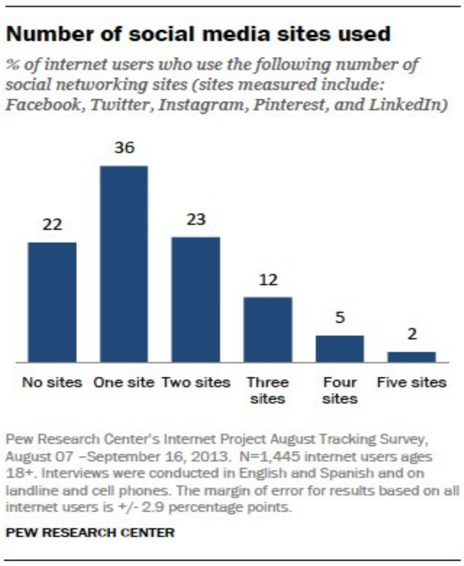
Who uses social networking sites

% of internet users within each group who use social networking sites

	All internet users	74%
a	Men	72
b	Women	76
a	18-29	89[cd]
b	30-49	82[cd]
c	50-64	65[d]
d	65+	49
a	High school grad or less	72
b	Some college	78
c	College+	73
a	Less than $30,000/yr	79
b	$30,000-$49,999	73
c	$50,000-$74,999	70
d	$75,000+	78

Pew Research Center's Internet Project January Omnibus Survey, January 23-26, 2014.
Note: Percentages marked with a superscript letter (e.g., [a]) indicate a statistically significant difference between that row and the row designated by that superscript letter, among categories of each demographic characteristic (e.g., age).
PEW RESEARCH CENTER

(a) Percentage of users that use the internet.



(b) Percentage of users that use multiple social networks.

Figure 1.2: Pew Research survey results show that out of 74 per cent users of internet, 42 per cent of them use Multiple Distributed (MuDi) social networks.

MuDi network users also cause the diverse set of activities that they perform becoming part of these networks. Some individuals behave in line with the standards and etiquettes of the specific network, while others intentionally fake identities to cheat others. Hence, a trust mechanism should draw information from all the MuDi networks involved before making any trust-related decision. This will not only base trust on diverse information but more accurately reflect their behaviour in multiple contexts.

There are different explicit and implicit set of activities that can become source of trust in online social networks (Zhang and Yu, 2012). Explicit actions happen as a result of direct actions of users for example, in friendship networks Facebook and Twitter, it can be a frequency of like/favourite or sharing/retweet. Similarly it can be an act of initiating friendship/follow relationship between users. Implicit activities on the other hand occur as a result of mutual activity, for example, in professional social networks, an act of collaboration/co-authorship frequency from networks can serve as a trust metric. Citation of someone else's work can also generate an implicit trust relationship due to being experts of the same research area. Both explicit and implicit types of trust values

[3]http://www.pewinternet.org/2013/12/30/social-media-matrix/

are different in nature and are based on varied set of activities but degree of involvement among users being part of the same environment can act a source of trust.

Unfortunately, consolidating multiple trust networks into a single network where trust calculations can be performed is non-trivial, as these are heterogeneous networks where the structure and weighting criteria are different. Care must be taken not to inflate or dampen trust values artificially. Not all the users are connected in all of the constituent networks, and in some cases will not even belong to some networks. Differentiating *absence of trust* from *distrust* is therefore a key issue with which any trust aggregation mechanism should be able to cope. But if the information and activities of users in the various social networks could be combined, it would provide a much richer dataset for making decisions about trust.

Linked data technologies are helpful when it comes to consolidating data from MuDi networks. The concept of linked data was proposed by Berners-Lee (2006) and compels data to follow a certain set of principles when creating links between different datasets. In consolidating MuDi social networks, it can help in linking multiple data repositories by identifying co-referred participants, linking them across multiple networks, thus generating a single global graph. There are already built-in ontologies and schemas that are stable and published, and they can be extended to model MuDi trust information. This procedure will generate a single network with multiple trust values between those who are connected by MuDi networks.

Aggregation of multiple trust values between co-referred users can be achieved using data fusion techniques. These integrate multi-source data into a single value, keeping in mind the reliability of the data, and the source of that data (Yager, 2004). The decision to specify reliability factor of different trust values and their sources is also challenging, as all networks are not of the same standard and they provide information that varies in quality and quantity. This emphasises the need to determine a method that can integrate information from MuDi networks while preserving the integrity of trust from each of the MuDi networks. If these methods can be efficiently applied then it would result in a single network with co-referred individuals with links between them representing aggregated trust measures.

Trust information between individuals with an interaction history is available from either some or all of the networks consolidated, but it needs to be evaluated for isolated users. Transitive decay-based trust propagation can undertake this task by evaluating trust for such users over the paths in social networks. Based on the studies by Holland and Leinhardt (1972) and Ziegler and Lausen (2004, 2005), researchers have proved that people prefer to trust a friend of a friend rather than a stranger, but that the strength of that trust weakens as the length of the path to the friend increases.

This thesis explores how to use data from MuDi networks for trust estimations and evaluates the approach using experiments on simulation and real world data. MuDi social networks were interlinked using linked data techniques, and multiple trust values were aggregated using different data fusion techniques. A simulation experiment generated pairs of networks with varying percentages of Participant Overlap (PO) and Tie Overlap (TO) to analyse the behaviour of data fusion techniques for a range of different networks. Results from the simulation showed that, of all the techniques, Weighted Ordered Weighted Averaging (WOWA) approach best consolidated and respected the integrity of trust from individual networks. Recommendations of the simulation experiment were then used for a real world experiment that extracted data from two professional social networks. Trust metrics from both individual and consolidated networks were compared with the real life trust values collected using a survey. Results of the real world experiment revealed that, for those users who are part of both the networks, consolidated trust metrics match real life trust better than metrics from either a cross-region (with one user present in both networks with the other in a single network) or from one of the individual networks.

1.2 Problem Statement

This thesis frames its work in terms of an expert recommendation scenario. Suppose that *Ben* and *Alice* are looking for a trusted person on the web in the area of *sociology* and *psychology*, *Bob* and *Charlie* and *David* are such persons. Imagine these users are part of two collaboration networks A and B (shown in Figure 1.3) where network A represents people working in the area of *sociology* while network B represents those working in *psychology*. Nodes in these networks represent researchers and weighted edges between them shows the trust that immediate neighbours on each edge holds about each other. There are three scenarios that can arise keeping in view the networks of *Ben*, *Alice*, *Bob*, *Charlie* and *David*.

1. *Ben* and *Bob* are members of both networks. Here, although they can approach each other in either of the networks, consolidation will permit them to find a trust path that aggregates trust information from both networks. For example, in Figure 1.3, *Ben* (Participant 2) finds *Bob* (Participant 3) as the right person, the consolidated network allows us to calculate aggregated trust between them from both the networks (for example, trust information on the link $2 \rightarrow 3$).

2. *Alice* and *Charlie* appear in the same single network. In this case, consolidated MuDi networks will aggregate trust metrics between overlapping users from different networks, which may effect the path between them. For example, in Figure 1.3 *Alice* (Participant 1) is searching for *Charlie* (Participant 4), and both are part of network A. Here, although they are connected to each other in the single network, the consolidated trust information between intermediate nodes (for example, $2 \rightarrow 3$, which appears in both networks) can impact on the trust value.

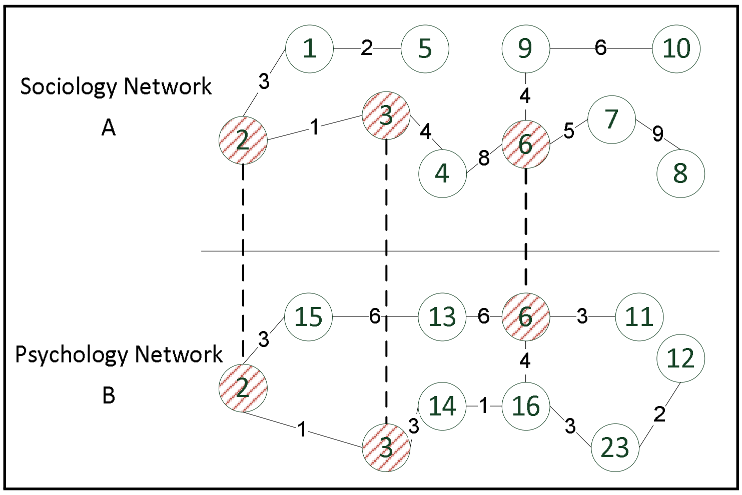


Figure 1.3: Sample networks (A and B) to describe the scenario of consolidating multiple social networks for trust computations. Pattern filled nodes represent participants who are part of both the networks.

3. *Alice* and *David* are part of different networks. Consolidated MuDi trust networks will enable *Alice* to locate *David* although they are present in different networks. In Figure 1.3, *Alice* (Participant 1) and *David* (Participant 15) are not in the same network, but consolidation has routed a trusted path to *David* through intermediate nodes (for example, $2 \rightarrow 15$).

1.3 Hypothesis

To test the potential benefits of consolidating MuDi social networks for trust aware decision making (described in Section 1.2), a set of testable claims are presented, termed H1, H2 and H3. This will also allow us to justify contributions of this research (mentioned in Section 1.5) by running experiments on both simulation and real world data.

H1 Semantic technologies allow us to *uniformly model* and annotate trust data from MuDi social networks for making trust computations over *heterogeneous resources*.

- **Heterogeneous resources:** When discussing in terms of the semantic web, each of the explicit and implicit sources of trust information has its own ontology for data representation. To make trust calculations over such a diverse set of resources, trust data needs to be mapped by a uniform representation (ontology).

- **Uniformly model:** Uniform model represents an ontology of trust that extracts trust data from variety of social networks on the web and represents that information in a single ontology for others to query and reuse that information.

H2 Data fusion techniques allow us to *aggregate trust metrics* from MuDi social networks and *respect the integrity of trust* from individual networks, while opening up many additional trust paths.

- **Aggregate trust metrics:** Multiple trust metrics between overlapping participants from MuDi social networks must be integrated to generate a single metric. This considers input from each of the metrics available from individual networks.
- **Respect integrity of trust:** The integrity of trust from individual networks must be respected by not inflating or dampening any of the individual trust data points.

H3 Trust metrics generated over MuDi social networks increase *accuracy of trust* over individual networks in terms of more accurately capturing how *users perceive* one another in real life.

- **User perception:** The trust that a user perceives in other people in the real life is the baseline, and is collected using a survey.
- **Accuracy of trust:** If the trust measurements from MuDi social networks approaches the users' perception more than those from individual networks, then we are justified to say that trust calculated over MuDi social networks improves the accuracy of trust metrics in comparison with those from single networks.

1.4 Methodology

The topic of research in this thesis covers multiple areas, including social networks, semantic web and trust, so the designed research methodology incorporates techniques from these different areas. This is to ensure that it thoroughly answers all the questions raised in the hypothesis section. This section works as an overview of the techniques discussed in later chapters.

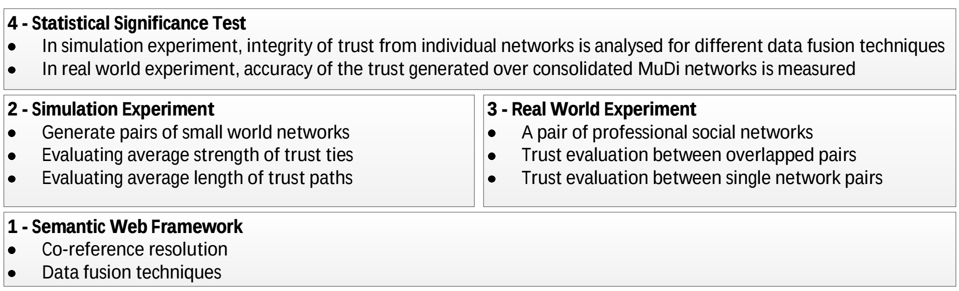


Figure 1.4: Component diagram representing the research methodology

The methodology of this work uses the proposed semantic web framework to run a simulation and a real world experiment. The aim of these is to evaluate the hypothetical claims of generating more accurate trust metrics using a consolidated version of MuDi social networks, compared to those from individual social networks.

1. **Semantic Web Framework** - The semantic web framework used the proposed OWL Lite trust ontology to annotate a simulation and a real world network data for trust annotations between participants in the network. It used a co-reference module to identify participants present in MuDi social networks, using the heuristic method of label comparison that matches the meta-data of users from multiple networks. A data fusion module aggregated overlapping trust metrics between co-referred users, and re-evaluated those available from any single social network. If the trust between any of the indirectly connected participants needed to be evaluated, then it used the principle of trust transitivity, and decay of trust existed along paths in the network.

2. **Simulation Experiment** - For the simulation experiment, the system first generated sample pairs of trust networks having varied percentages of Participant Overlap (PO) and Tie Overlap (TO). The generated networks were then consolidated using the proposed semantic web framework. This annotated and applied all the operations to aggregate or re-evaluate trust metrics between participants with the aim of respecting the integrity of trust data from both the sample social networks. The average strength of trust ties (TS) and average length of trust paths (TL) in the respective networks were calculated using various data fusion techniques, and Weighted Ordered Weighted Averaging (WOWA) emerged as the one that satisfied the condition of trust preservance. WOWA was then recommended for use in consolidating the real world networks.

3. **Real World Experiment** - The real world experiment selected a pair of professional social networks, WAIS (Web and Internet Science) ePrints co-authorship [4] and WAIS projects collaboration networks [5], to test the accuracy of trust metrics generated over consolidated pair of networks. The proposed semantic web framework generated the resultant single consolidated network by finding the co-referred users who are in both networks. Already selected data fusion technique, WOWA was used to aggregate the overlapping trust metrics and to re-evaluate those available from cross-regions or only one of the networks with the aim of analysing the accuracy of consolidated trust from different types of participant pairs.

4. **Statistical Significance Test** - Statistical significance of the results in both the simulation and the real world experiment was evaluated by applying the T-Test and

[4] co-authorship network among researchers publishing research articles in WAIS research group (extracted from ePrints publication network), http://www.wais.ecs.soton.ac.uk/publications

[5] collaboration network among researchers working on funded projects in WAIS research group (extracted from WAIS projects network), http://www.wais.ecs.soton.ac.uk/projects

the generated p-value was analysed. For the simulation experiment, the p-value was analysed to prove whether the claim of generating trust metrics that respect the integrity of trust from individual networks was justified. In the case of real world networks, the p-value tested the claim of whether the consolidation of MuDi networks improved the accuracy of trust metrics compared to those generated by individual social networks.

1.5 Contributions

The research work completed in this thesis contributed the following novel theories to the existing work in the area of trust and the semantic web:

- A semantic web framework is proposed that consolidates MuDi social networks and generates aggregated trust metrics between users publishing resultant data as linked data.
- A trust aggregation scheme is proposed that considers the importance of trust data and trust data source, and respects the integrity of trust data from individual social networks.
- It is demonstrated that existing mechanisms of trust evaluation on single social networks do not reflect real world views so well as trust values calculated from MuDi networks. In addition, trust evaluation mechanisms of a single social network do not allow trust calculations about those in other social networks; hence the idea of MuDi social networks can allow us to go beyond the boundaries of individual networks.

One conference paper carrying simulation work has been accepted and published in ASE Human Journal Imran et al. (2012).

1.6 Thesis Structure

The thesis comprises seven chapters in total, each covering one of the key components of the whole.

Chapter 1 introduces the idea of trust evaluations over MuDi social networks and explains the motivation behind the proposed idea.

Chapter 2 gives background material relating to trust in social networks and discusses existing tools and techniques. This includes different types of social networks on the web that can contribute trust information, data fusion techniques that allow us to consolidate multiple trust values, and the affordance of semantic Web technologies in this regard.

Chapter 3 proposes a semantic web framework for interlinking MuDi social networks to generate a single graph that represents a consolidated version of many individual social networks. Different trust annotation methods are discussed and trust ontology is proposed that allows us to annotate trust data.

Chapter 4 presents different data fusion techniques for trust aggregation and discusses the advantages and shortcomings of each. Furthermore, it discusses two trust propagation algorithms for evaluating trust for distant participants using the trust transitivity principle.

Chapter 5 runs the simulation experiment by consolidating two randomly generated social networks having varying percentages of PO and TO. Results obtained by using different data fusion techniques and different trust propagation algorithms are analysed for preserving the integrity of trust from individual social networks.

Chapter 6 runs a real world experiment by consolidating two professional social networks and a trust survey to collect actual proxy trust values between participants in real life. Results from the consolidated version of networks are analysed for accuracy in comparison with individual social networks for *overlapping*, *cross-region* and *single-network* pairs of participants.

Chapter 7 discusses the results from Chapter 3, 5 and 6 with reference to each of the hypothesis statements. Furthermore, it concludes the thesis by listing what has been achieved and by describing future research directions emerging from this work.

1.7 Summary

This chapter provides an overview of the work and describes the motivation behind using MuDi networks for trust calculations. It is emphasised that the presence of multiple social networks, each with different functionality, creates the opportunity to explore a variety of data for trust evaluations. Keeping this in mind, real life scenarios are presented that better situate the work in the real world on improving existing systems. Next, it describes three testable claims that allow us to assess the merit of the proposed idea and then elaborates the methodology for testing those claims. Following, it lists the contributions of this research in the area of trust and the semantic web, and finally presents the structure of this thesis.

The next chapter gives the background of the work and describes existing techniques that can act as a guidelines for developing a MuDi solution.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews relevant studies in the literature about trust in social networks. First, it gives a multidisciplinary view of trust and includes definitions from psychology, sociology and social networks. Then, it defines explicit and implicit types of online social networks along with their characteristics. Different data fusion techniques for aggregating multiple trust metrics available between users are discussed. The affordances of semantic technologies with respect to trust data modelling and management from multiple networks are also reviewed. This chapter concludes by giving the specific definition of trust used in this thesis.

2.2 Trust in Online Social Networks

Trust is the foundation of all human interactions, either on the web or in real life, but in online social networks it is especially important because people interact potentially anonymously. Malicious users leave other users in the network vulnerable to their activities and a concept of trust may minimise the risks of being cheated.

2.2.1 Defining Trust

Trust is a multidisciplinary concept and many authors have proposed definitions. Rotter (1971) describes it as a psychological phenomenon: 'trust is *the generalized expectancy that the statements of others can be relied on or promises will be fulfilled*'. According to Rousseau et al. (1998) it is '*a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behaviours of another*'. Cook and Wall (1980) define it as '*the extent to which one is willing to ascribe good* *intentions to and have confidence in the words and actions of other people*'. Fukuyama (1995) believes that trust creates social capital (Coleman, 1988) and makes an enviroment for people to work and collaborate with each other; it is '*the belief, or willingness to believe that one can rely on the fairness, goodness, strength, and ability of somebody*'.

Online trust, and specifically in social networks, is either based on direct set of experiences between people or as an inferred metric evaluated from the experiences of others using trust recommendation mechanisms (Corritore et al., 2003). Furthermore, it can be based on reputation information that calculates global value (giving a value for each individual in the network; Resnick et al. (2006)), or can be calculated locally (giving multiple values, depending on which node you choose to look from; Golbeck and Hendler (2006a)).

Xiong and Liu (2004) define reputation-derived trust as '*an evaluation it receives in providing services to other peers in the past*'. Wang and Vassileva (2003) state that trust is a personalised metric: '*it is a peer's belief in another peer's capabilities, honesty and reliability based on its own direct experiences*' . Golbeck and Parsia (2006) considers trust as a matter of personal commitment, whereby '*trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome*'. Walter et al. (2008) describe indirect trust over trust paths: '*trust is an expectancy of an agent to be able to rely on some other agent's recommendations*'.

Some researchers do believe that the reputation and trust are two different concepts, and a person carrying bad reputation does not always imply distrust. According to Mui et al. (2002) trust is '*a subjective expectation an agent has about another's future behaviour based on the history of their encounters*' while reputation is '*a perception that an agent creates through past actions about its intentions and norms*'. For this work, it is assumed that the reputation helps establishing trust and carries the information that builds image of the person in the eyes of others.

Another aspect of trust is its subjectivity, in that it does not hold in all areas equally; rather, it is a subjective value that represents an individual's belief in that relevant area. Grandison and Sloman (2000) describe it as '*the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context*'. Olmedilla et al. (2006) define the '*trust of party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period (within a specified context in relation to service X)* .

This work deals with the *personalised subjective* trust that exists either between users as a result of direct interactions or is inferred, based on the recommendation of other trusted friends in the network. Further discussion about how it can be mapped in the context of multiple social networks is described in Section 2.8.

2.2.2 Trust Metrics Classification

The trust metric expresses a quantitative estimate of the trust that two participants in the network hold about each other. As mentioned in Section 2.2.1, it can either use direct experiences or the recommendations of trusted friends. The evaluated trust metric can be broadly categorised as global and local. Global trust metrics compute reputation values (the work of Pagerank is a notable example: see Lawrence et al. (1999)). It considers whole-network information and results in each node of the network receiving a single objective trust value. Local trust metrics are based on calculations from a given individual's position in the network, so each node has its own subjective view of the trust of every other node (Mui et al., 2002).

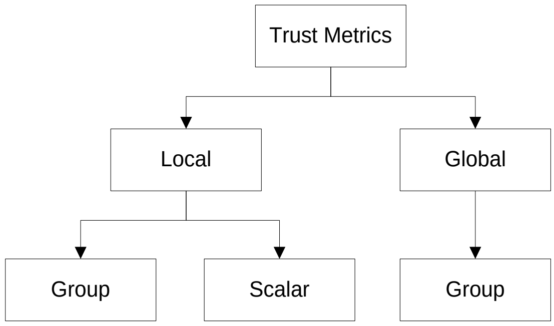


Figure 2.1: Classification of Trust Metrics taken from Ziegler and Lausen (2004).

Trust metrics can be further categorised as scalar and group; as described by Ziegler and Lausen (2005). Scalar metrics deal with individual assertions for evaluation, such as seeking recommendation from only one of the friends, while group metrics consider a group of assertions for that purpose, such as evaluations based on the recommendations of all the friends. If these categories are coupled with the local and group categories described above then it gives a clearer picture; scalar metrics are always local, because they search for a highly trusted path between participants (Walter et al., 2008), while group metrics can be either local or global. Local group metrics aggregate information from all the paths leading from source to destination (Levien, 2009; Ziegler and Lausen, 2004), and global group metrics need network-wide information (Zhang et al., 2006).

In this work we are concerned with a *scalar local* trust (as online trust tends to be personalised and subjective (Golbeck and Hendler, 2006b)). The trust calculation starts by taking source to destination participants as input parameters then calculates the trust of the destination using a trusted path (selection of the path is based on trust algorithm discussed in Section 4.7).

2.2.3 Trust Evaluation Techniques

As explained in Section 2.2.2, interpersonal trust in online social networks can be viewed either as a global value built on the reputation of the person in the community or some local value pertinent to direct interactions or interaction(s) with other trusted friend(s) in the network. There are different ways of generating reputation value; first by analysing explicit trust ratings about people, for example, the reputation of a seller in eBay; second, by analysing links with other people in the network, such as the way Google ranks web pages based on the links with other web pages. Local trust estimation involves finding trust path(s) through directly connected trusted friend(s) using trust propagation mechanisms.

There are many articles in the literature that summarise these different evaluation techniques. For example, Sherchan et al. (2013) provides a comprehensive overview of the social trust and discusses three aspects of it: *trust information collection*, *trust evaluation*, and *trust dissemination*. Further, it summarises already proposed techniques related to each of these aspects. Moyano et al. (2012) proposes a conceptual framework that compares and analyses different trust and reputation models. It classifies them into two categories, decision models and evaluation models. Decision models include policies and negotiation strategies that use negotiation-driven exchange of credentials and policies before establishing trust between any two parties. On the other hand, evaluation models use behaviours of the people for establishing trust between users. For example, it can either be a set of past experiences or based on interactions of others using propagation mechanisms.

2.2.3.1 Reputation-based Trust

One of the applications of reputation-based trust is the generation of global trust values for sellers in the eBay system. From the three-scale rating (positive, negative, neutral), a net reputation score for each seller is calculated by subtracting distinct negative ratings from distinct positive ratings. Textual feedback is also displayed, with the most recent comment on the top for customer to read before buying anything. Pagerank also gives a perspective of trust and generates a single objective value about each participant in the network. Mtibaa et al. (2010) proposed pagerank-based algorithm called *PeopleRank* to rank people for data delivery in an opportunistic social network. Like the pagerank algorithm used to crawl the web to rank pages based on their relative importance, people in a social network are ranked according to their connection with other trusted people in the network.

$$PeR(N_i) = (1-d) + d \sum_{N_j \in F(N_i)} \frac{PeR(N_j)}{\mid F(N_j) \mid} \tag{2.1}$$

where $N_1, N_2, \cdots, N_n$ are the nodes representing people in the network, $F(N_i)$ is the set of neighbours connected with N_i, and d is the dampening factor that defines the probability that the social relation can improve the rank of connected people.

Nepal et al. (2011) uses social capital to build trust communities and proposes a social trust model, *STrust*. It is based on two sets of information; first is the multi-context popularity trust, which evaluates overall reputation of the user in the network. Second is the engagement trust which assesses the participation of the person in the network. Equation 2.2 shows the procedure.

$$STrust(u_i) = \alpha.PopTrust(u_i) + (1 - \alpha).EngTrust(u_i) \tag{2.2}$$

where α represents the weight in the range 0 to 1. It indicates the reputation of the participant in the network. The sustainability of this model with respect to social capital was conducted using a recommendation application (Nepal et al., 2013).

The main drawback of reputation-based global trust is the risk of exploitation due to altruistic behaviour of certain participants, as mentioned by Resnick et al. (2006) for eBay. It happens when fake, selfless nodes are injected into the network to afford a high trust rating to certain nodes, in order to increase their good reputation for others to trust. This results in people trusting that node and they are ultimately cheated.

2.2.3.2 Peer-to-peer Based Trust

Peer-to-peer models evaluate local values of trust, either based on direct interactions or on the basis of recommendations from trusted peers in the network. The key difference between these models and the centralised eBay reputation is its decentralised management of trust, as it is not possible for a member to have information about all other members in the network.

EigenTrust, proposed by Kamvar et al. (2003), defines trust as a function of corrupt versus valid files provided by the peer in the peer-to-peer file sharing environment. Each member maintains history of interactions with peers, based on the files received from them. Trust is equal to the difference of satisfactory ($sat(i,j)$) and unsatisfactory ($unsat(i,j)$) file downloads.

$$s_{ij} = sat(i,j) - unsat(i,j) \tag{2.3}$$

If there is no direct information available, a member asks other peers for their local trust values. These local trust values are normalised and a final trust value c_{ij} is calculated using Equation 2.4.

$$c_{ij} = \frac{max(s_{ij}, 0)}{\sum_j max(s_{ij}, 0)} \tag{2.4}$$

Griffiths et al. (2006) proposes a fuzzy trust model that calculates subjective trust values, based on the person's direct experience of the peer. Interaction history relating to each aspect of trust is maintained separately, and four different levels of trust are specified using fuzzy logic. Equation 2.7 shows the method of calculating *experience*, E^d_α based on the history of good and bad interactions.

$$E^d_\alpha = \frac{I^{d+}_\alpha - I^{d-}_\alpha}{I^{d+}_\alpha + I^{d-}_\alpha} \tag{2.5}$$

where I^{d+}_α represents the number of good interactions and I^{d-}_α shows the number of bad interactions.

Wang and Vassileva (2003, 2005) propose a Bayesian network-based model for specifying multi-faceted trust in a file sharing environment. The three considered trust aspects are the *type*, *quality* and *download speed* of the file. A source agent can consider any or all of these parameters to measure the file provider's capability. Equation 2.6 represents the probability relationship between different elements of the model,

$$p(h \mid e) = \frac{p(h \mid e).p(h)}{p(e)} \tag{2.6}$$

where $p(h)$ is the prior probability of the hypothesis h to be analysed, $p(e)$ is the prior probability of evidence, $p(h \mid e)$ is the probability of h, given e, and $p(e \mid h)$ is the probability of e, given h.

2.2.3.3 Transitive Decay-based Trust

Decay-based models in the literature are mainly used for calculating local subjective trust values in online social networks and are much like peer-to-peer techniques, but with trust decay (Guha, 2003; Josang et al., 2003) over paths in social networks, missing from peer-to-peer systems.

Trust decay is based on the transitivity principle, first discussed by a psychologist, Heider (1958), in the context of the transitivity of positive interpersonal sentiments. Later on, social psychologists, Holland and Leinhardt (1972) examined this in terms of social relations by running an experiment over a set of 917 sociograms asking people about their sentiments relating to other people in the group. They found that in 70 per cent of the groups there was a significant tendency towards transitivity. In another study by Leinhardt (1972), 118 sociograms of children of different age groups were analysed for transitivity of positive sentiments; results showed an increasing tendency towards transitivity with an increase in age.

Trust transitivity in social networks allows people to trust friends of their friends over trust paths, but the strength of that trust weakens as the length of path to friend increases (Guha et al., 2004; Josang et al., 2003). For example, taking Figure 2.2 as an example, although there exists no direct connection between *Alice* and *David*, because they are connected through *Bob* and *Charlie*, there is a *weakened trust* between them propagated through directly connected participants.

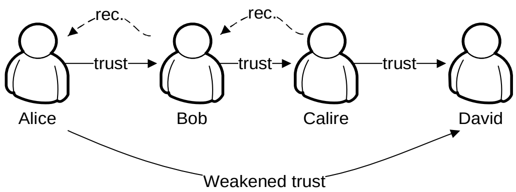


Figure 2.2: Trust transitivity between users *Alice* and *David* (taken from (Josang et al., 2003)).

There are a number of trust evaluation techniques using the concept of trust decay, for example, Ziegler and Lausen (2005) describe a trust and distrust propagation mechanism, '*Appleseed*', using a spreading activation method (Ziegler and Lausen, 2004). Trusted friends pointing to the person in graph inject energy that is distributed along all the paths originating from that person. Based on the empirical evidence, a realistic decay factor is chosen and trust propagates to each of the neighbour nodes, with a normalised local edge weight, $e_{x \to y}$, assigned to each edge.

$$e_{x \to y} = d.in(x).\frac{W(x,y)^2}{\sum_{(x,s) \in E} W(x,s)^2} \tag{2.7}$$

where d specifies the decay, *n(x)* represents the amount of trust injected into node x, *W(x,y)* represents the weight of the link between nodes x and y and *W(x,s)* represents the weights of all outgoing edges used for normalising the trust.

Golbeck (2005) trust inference algorithm searches for trust paths moving from source and destination in the forward direction, and then multiplies trust values available on those links (in the range [0,1]) moving backward towards the source, thus called the *TidalTrust* mechanism (Golbeck and Hendler, 2006b). The weighted average of the values is taken to calculate the final trust value if there are multiple paths.

$$r_{is} = \frac{\sum_{j \in adj(i)} t_{ij} t_{js}}{\sum_{j \in adj(i)} t_{ij}} \tag{2.8}$$

'*FilmTrust*' is an application of the *TidalTrust* inference scheme that recommends movies to users of a network on the basis of similarity in their movie taste (Golbeck and Hendler, 2006a). Users share movies in the network and other users in the network specify their level of trust in that user, keeping in mind their like/dislike of that movie. Ratings of the trusted users are normalised to find the recommended rating r_{sm}.

$$r_{sm} = \frac{\sum_{i \in S} t_{si} r_{im}}{\sum_{i \in S} t_{si}} \tag{2.9}$$

Lesani and Bagheri (2006) improve on *TidalTrust* algorithm in two ways; first, they translate trust as a fuzzy set (for example, low, medium, high, etc) rather than a scale (0, 2.5, 10, etc), claiming that it is more meaningful for users, and second, they not only consider the shortest paths as trusted paths, believing in stronger paths that may be longer than the shortest. The algorithm sets an initial threshold and only participants having trust greater than this threshold are selected as trustworthy members.

Walter et al. (2008) presents a trust-based recommendation system, part of which is to design a propagation mechanism for inferring trust to distant nodes. It takes trust transitivity as the basis and multiplies trust information (in the range [0,1]) available on the link over the path to arrive at the final trust value. Suppose the trust path between participants a_i and a_j is $path(a_i, a_j)$, then the Equation 2.10 is used for evaluation of trust,

$$T_{a_i,\ldots,a_j} = \prod_{(a_k,a_l) \in path(a_i,a_j)} T^{a_k,a_l} \tag{2.10}$$

Liu et al. (2010) also calculate trust using the multiplication of trust values between adjacent participants over the trust path.

$$T_p a_1, \ldots, a_n = \prod_{(a_i,a_{i+1}) \in P(a_1,\cdots a_n)} T_{a_i,a_{i+1}} \tag{2.11}$$

Unlike Ziegler and Lausen (2005) decay of trust in Walter et al. (2008), Golbeck and Hendler (2006b) and Liu et al. (2010) is not controlled by the source. Instead, discounting of trust happens as a result of multiplication of individual values along the path.

Kim and Song (2011) predicts local trust in social networks using the concept of reinforcement learning. Trust is propagated towards the destination using different strategies with the aim to select the strategy that best predicts with the maximum accuracy. Results showed that the hybrid approach of min-max and weighted mean aggregation over all available shortest paths turned out to be the best approach.

Similarly, Verbiest et al. (2012) presents an idea of incorporating trust paths of different lengths for enchancing accuracy of trust calculations. Idea of trust decay is considered and metrics are analysed for both shortest and longest trust paths. Results showed that longer paths have more errors in their evaluations than shorter paths.

Another research conducted by Chakraborty and Karform (2012) compares three different versions of the trust propagation algorihms with the one already developed known as *MoleTrust* (Avesani et al., 2005). Results showed that decay-based trust algorithm outperformed other two and the *MoleTrust* algorithm.

This study adapts the simplest of all the techniques, presented by Walter et al. (2008) and Liu et al. (2010) as the starting point to implement and analyse the consolidation of trust over multiple networks. These algorithms are analysed against the strongest and shortest trust paths for a set of derived trust factors between users in the network. Further discussion is in Section 4.7.

2.3 Online Social Networks

Attempts at consolidating multiple social networks have been made in studies reported in the literature, but none have investigated the impact on trust-related measures. This work specifically targets this area and analyses the aggregated trust metrics from these networks. But before incorporating multiple social networks into trust measurement, it is important to understand the different types of trust data that these networks provide and the characteristics of these networks.

2.3.1 Categorising Social Networks

Online social networks can be viewed as explicit and implicit types of networks; explicit networks are formed due to deliberate actions of users, for example, to add someone as a friend, rating a seller and so on. These networks reflect a user's choice and, although representing a user's direct trust, provide little support for discovering new information. Implicit networks, on the other hand, are bipartite networks that emerge as a result of mutual activities of users in the shared environment, such as participation in online forums, co-authoring publications and so on. This type of network helps in extracting new knowledge and discovering new relationships, lacking in explicit networks. For example, Hwang et al. (2010) uses implicit co-authorship networks for literature recommendations; this method uses a set of recently accessed articles by the researcher and recommends similar articles by direct or indirect co-authors in the network.

Figure 2.3 shows both type of networks, with explicit networks giving more clear knowledge, with trusted users having links between them, while implicit networks represent proxy trust information that can be extracted by linking users to each other (as in explicit networks) due to commonality of actions. Many researchers have categorised these networks for the functionality they provide; for example, Newman (2001a) and Golbeck (2006) group them as blogging, business, dating networks and so on.

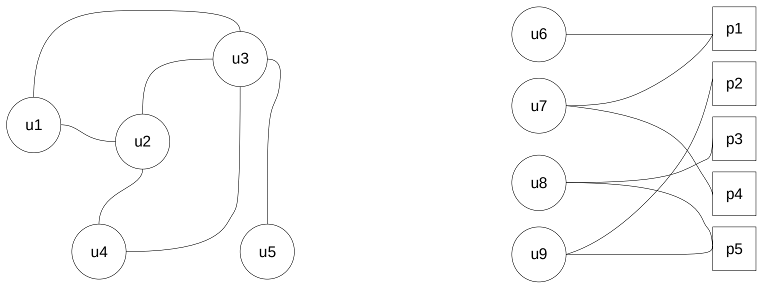


Figure 2.3: Explicit and implicit networks (left and right respectively), with u showing *user* and p representing research *article* published by researchers.

Both explicit and implicit networks are believed to be small-world networks (Amaral et al., 2000), for example, Mislove et al. (2007) perform structural analysis on different explicit social networks, measuring indegree, outdegree, clustering coefficient and other properties, and find that these networks have small-world and scale-free properties. Another study conducted by Ahn et al. (2007) also proves that real-world social networks are small-world networks.

2.3.2 Characteristics of Online Social Networks

The two properties characterising small-world behaviour of online social networks are clustering coefficient (represented using C) and shortest path length (mentioned as L) (Watts and Strogatz, 1998). C represents the level of clustering in the network and its value ranges from 0 to 1, scaled from low to high. It measures the extent to which nodes in the network are connected to each other and ensures transitivity as, in most cases in social networks, two friends of a single person are also friends of each other. The value of C for the network can be calculated using Equation 2.12:

$$C = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{2.12}$$

where n is number of users in the network and c_i represents local clustering of each user and its value for an undirected network can be calculated as shown in Equation 2.13.

$$c_i = \frac{e_i}{\frac{1}{2}k_i(k_i - 1)} \tag{2.13}$$

where e_i represents actual number of ties and $k_i(k_i - 1)/2$ is the maximum possible number of ties between neighbours of user i. The other small-world property, L, is the length of the shortest path between pairs of participants in the social network and its average value for the undirected network can be calculated using Equation 2.14 (Barrat and Weigt, 2000).

$$L = \frac{1}{\frac{1}{2}n(n-1)} \sum_{s,t\epsilon N} d(s,t) \tag{2.14}$$

where N represents set of n users and *d(s,t)* is the length of shortest path from s to t.

Watts and Strogatz (1998) describes the small-world networks with respect to the randomness as depicted in Figure 2.4. It states that these networks reside between regular and random networks. Like regular networks they have high value of C and similar to the random networks contains short connections which results in low value of L. The two step procedure for creating small-world network is described below.

1. First, set of nodes N in the network is connected in the form of a ring lattice. This is to ensure that each node has k number of links which helps in establishing connectedness of the network.

2. In the next step, randomness is introduced by rewiring links between uniformly selecting random nodes with probability r unlike the first case where nodes were connected in a predetermined fashion. Duplicate edges and self-loops are forbidden to ensure same number of links as were in the first step.

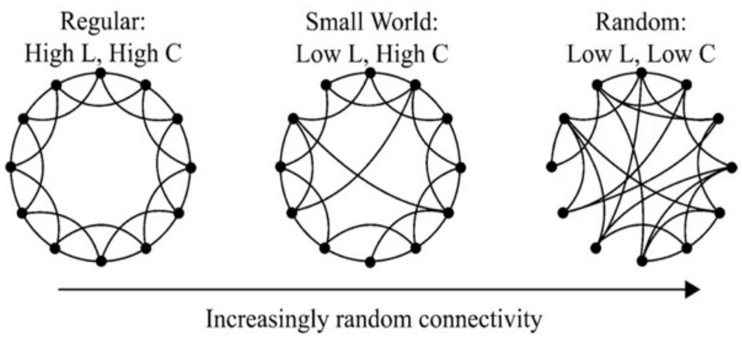


Figure 2.4: Transition from a completely regular network to a random network with respect to the randomness factor r. Small-world network lies in-between having high value of C and low value of L.

This method has been extended in Section 5.2.1 for generating small-world networks but there it considers an additional factor of overlapping/ non-overlapping nodes and links.

Apart from the small-world properties, another parameter that impacts on the structure of network is the density (indicated by D). While generating a network, certain value of C and L in the network can only be ensured if the network has a certain value of D, because otherwise it will end up having a deficiency of connections. The value of D in any social network is the ratio of the number of connections in the network to the total number of possible connections, assuming there are no self-loops in the network. For a network of n users, network density for an undirected network can be calculated using Equation 2.15:

$$D = \frac{\#ofties}{\frac{1}{2}n(n-1)} \tag{2.15}$$

Table 2.1 shows values of these parameters for several co-authorship networks collected from the literature (Newman, 2001b,a). These values act as a guideline when generating random networks for consolidating multiple networks in the simulation experiment (Section 5.2.2).

Table 2.1: Small-world properties of co-authorship networks taken from (Newman, 2001b,a).

Co-authorship Networks	No of Nodes	C	L	D
Physics	52909	0.56	6.19	0.03504
Biology	1520251	0.6	4.92	0.00204
Math	253339	0.34	7.57	0.00309

2.3.3 Consolidating Multiple Networks

There are a range of explicit and implicit social networks on the web, but existing trust mechanisms rely on a single network for evaluation of trust. Aggregation of these different types of networks could allow us to base trust on the different types of relationships between individuals. The idea of consolidating MuDi social networks has its application in many other areas, including representation of separate social networks in a single social graph and processing of distributed data for reuse in search applications. They are, however, limited either to combining networks or to using their data in search systems, and the impact of different consolidation parameters on trust factors remains unexplored, to date.

There are many examples of researchers combining social networks. For example, Tang et al. (2008) propose a search system for academic researchers using semantic technologies. Their system uses Google API for extracting and integrating information about researchers from distributed locations on the web. Guy et al. (2008) propose *SONAR*, an API that can integrate information about users from multiple social networks; they claim that it can give complete and useful picture for end users. To use this system, however, API needs to be installed on all systems from where data need to be extracted. Similarly, another system, *Polyphonet* extracts and analyses network information from multiple social networks (Matsuo et al., 2006). Integrated information is analysed to determine, for example, degree distributions and path length.

Bae and Kim (2009) work integrates separate social networks into a single global social graph for analysis, using the concept of a hypergraph. Resources in multiple networks are connected using hyperedges and connections between them are normalised to depict a single social network. However, their work examines the resulting hyper-structures rather than attempting to consolidate weights or tie meta-data.

Similarly, the system proposed by Mika (2005), *Flink*, helps in the extraction, aggregation and visualisation of social networks. It extracts semantic data from web pages, emails and publication archives into the 'Sesame' (a triplestore) then uses semantic web technologies to aggregate that data.

2.3.4 Semantic Social Networks

Since Berners-Lee et al. (2001) proposed the idea of a semantic web, organisations and social networks (both explicit and implicit) have started putting their data on the web (under open data initiative [1]), mostly with clearly defined ontologies. These are usually built by reusing already defined ontologies (for example, FOAF (Brickley and Miller, 2010) is used for representing networks of people) and the decision to extract semantic data from such networks provides uniformly modelled semantic data (Jung and Euzenat, 2007). This eliminates the need to pre-process data from a variety of sources (for example, NodeXL, pajek, etc) to make them compatible with each other before running any specialised algorithms.

There are many deployed social networks which allow us to crawl semantic data using APIs provided by them; for example, flickr data can be accessed using flickrAPI and there is already one sample built over that (flickr^{TM} wrappr [2].) which extends dbpedia with links to the Flickr photos. Similarly twitterAPI can be used to extract twitter data about users and the tweets. Facebook has started providing access to semantic data of its users under the Facebook Graph API (Weaver and Tarjan, 2013).

One type of semantic social networks is collaboration networks (a type of implicit network) and there are many such networks for representing RDF information about articles published by researchers. For example, ePrints is a publication dataset that represents information about research articles published by people working at the University of Southampton. It uses vocabularies such as Dublin Core (Board, 2012) and FOAF (Brickley and Miller, 2010) that presents information about users and articles published by them. Similarly, there is another ontology designed here in the Electronics and Computer Science (ECS) department (reusing existing ontologies) at University of Southampton,

[1] http://globalopendatainitiative.org/
[2] http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/

known as ECS ontology (ECS, 2013), for semanticising data about those working in the school and collaborating on different projects.

2.4 Data Fusion Techniques

Consolidating MuDi networks is essentially a data fusion issue and more specifically one of data aggregation. A simple form of aggregation is to use Summation (S) or an Average (A) of the values. This works for simple numerical information (such as counts), but in the case of trust these could damage the trust values by either inflating them (through summation) or dampening them (when averaging several weights, some of which may be effectively zero as they are missing from one of the networks being aggregated).

There are two parameters that need to be considered for trust aggregation, **1)** reliability of information and **2)** reliability of the source of that information. A number of data fusion techniques exist in the literature that may be used for this purpose (Vicenc and Yasuo, 2007a,b; Xu and Da, 2003; Yager and Kacprzyk, 1997). Some of these techniques consider one of these parameters, while others consider both. It is important to note that the terms of reliability, importance and weight are used interchangeably in this thesis.

2.4.1 Weighted Averaging

An alternative method for aggregating trust data that better protects the values being aggregated is to consider only the reliability of the sources of information by taking the Weighted Average (WA). The reliability of the source is represented by the weight vector p and that is the only parameter input into the system, other than the trust data. A mapping function f_{WA}: $\mathbb{R}^n \rightarrow \mathbb{R}$ is a WA operator of dimension n if:

$$f_{WA}(a_1, a_2, \ldots, a_n) = \sum_i p_i a_{(i)} \tag{2.16}$$

where weight of each source is non-negative and $\sum_i p_i = 1$.

Although this method captures the source reliability, it misses the importance factor for the data provided by that source. As a result, it can severely damage the integrity of information by treating missing data in the same way as data with the value zero.

2.4.2 Ordered Weighted Averaging

Yager (1988) proposed an Ordered Weighted Averaging (OWA), an aggregation operator that considers the relative importance of the information used for aggregation. Unlike simple operators, it prioritises values in descending order, and allows us to assign weights bearing in mind the position of the information in the ordering. Here, the weight vector w is used to represent the importance of the trust data. A mapping function f_{OWA}: $\mathbb{R}^n \rightarrow \mathbb{R}$ is a OWA operator of dimension n if:

$$f_{OWA}(a_1, a_2, \ldots, a_n) = \sum_i w_i a_{\sigma(i)} \tag{2.17}$$

where σ is permutation that orders the elements: $a_\sigma(1) \leq a_\sigma(2) \leq \ldots \leq a_\sigma(n)$. All the weights are non-negative values and their sum should be 1, i.e. $\sum_i \omega_i = 1$

Unlike WA, OWA allows us to preserve the integrity of trust values from individual networks. The weight vector, w, serves that purpose by assigning high weights to high trust values than lower trust values.

Although this mechanism is claimed to be better than WA as it respects the integrity of each data point out of multiple ones, it ignores the importance of the source of that information, hence still provides an incomplete picture.

2.4.3 Weighted Ordered Weighted Averaging

The disadvantages of both the WA and OWA are eradicated in the Weighted Ordered Weighted Averaging (WOWA) technique proposed by Vicenc (1997), that considers the relative importance of both trust data and its source.

The importance of the information and its source are represented using two weight vectors, w and p respectively, each of dimension n. $p = [w_1, w_2, \ldots, w_n]$ and $p = [p_1, p_2, \ldots, p_n]$ such that i) $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$ ii) $p_i \in [0, 1]$ and $\Sigma_i p_i = 1$. A mapping function f_{WOWA}: $\mathbb{R}^n \rightarrow \mathbb{R}$ is a WOWA operator of dimension n if:

$$f_{WOWA}(a_1, a_2, \ldots, a_n) = \sum_i \omega_i a_{\sigma(i)} \tag{2.18}$$

where $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}$ is an ordering of n trust values, such that $a_{\sigma(i-1)} \geq a_{\sigma i}$, $i = 2, 3, \ldots, n$. Weight ω_i for the i^{th} data point in Equation 2.18 is defined as:

$$\omega_i = w^* \left(\sum_{j \leq i} p_{\sigma(j)} \right) - w^* \left(\sum_{j < i} p_{\sigma(j)} \right) \tag{2.19}$$

where w^* is a monotonic function (e.g., a polynomial) that interpolates the points $(i/n, \Sigma_{j \leq i} w_j)$ along with point $(0, 0)$.

Unlike WA and OWA, WOWA contains two weight vectors w and p. In context of trust, w is used to respect the integrity of high trust values while p allows us to rank sources of high trust values as high. This method fulfils both the conditions of considering importance of both the individual trust values and their values and hence generates trust values that according to our claim are better in quality. Further discussion about this claim and its accuracy is presented in Chapter 6 using a real-world experiment.

2.4.4 Induced Ordered Weighted Averaging

There are situations when trust data from certain sources are more worthy than those from others. Data from such sources must be weighted high and Induced OWA (IOWA) proposed by Yager and Filev (1998, 1999) allows us to do this.

The IOWA operator specifies data to be aggregated along with order of each data point using a tuple $<p_i, a_i>$, where p_i is the *order inducing value* and a_i is the *trust value*. The weight vector w specifies the importance of the corresponding ordered induced argument values. So there are two reliability vectors in this aggregation mechanism, other than the trust data **1)** $p = \{p_1, p_2, \dots, p_n\}$ **2)** $w = \{w_1, w_2, \dots, w_n\}$ where $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$. A mapping function f_{IOWA}: $\mathbb{R}^n \rightarrow \mathbb{R}$ is a IOWA operator of dimension n if:

$$f_{IOWA}(<p_1, a_1>, <p_2, a_2>, \dots, <p_n, a_n>) = \sum_i w_i a_{p(j))} \tag{2.20}$$

where p is the index function such that $p(j)$ is the index of the argument pair with the j^{th} largest order inducing value. Like the WOWA, IOWA allows us to set the weight of both trust data and their sources with vectors w and p respectively. But unlike the former one, it prioritises trust data with respect to sources that provide data, and not in the descending order, which was happening in OWA and WOWA.

All these data fusion techniques are analysed and tested for trust aggregation using a simulation experiment (in Section 4.5). In this work, to ensure the integrity of trust values, weights of high trust values and their sources are assumed to be high compared to lower trust values. However, weights can also be learned from data as discussed by Filev and Yager (1994); Yager (2003).

Most recently, there has been discussion about using data fusion techniques for aggregating multiple trust scores. For example, the work by Victor et al. (2011) aggregates trust scores from multiple trust paths between any two users in the network using different aggregation techniques. Results showed that K-OWA (knowledge awarding-OWA) and KAAV (knowledge awarding-averaging) performed better than other techniques. Similarly Ma et al. (2014) used aggregation techniques for generating realistic trust scores over transitive triads. Results showed improved performance than already existing techniques. The work by Bistarelli and Santini (2014) considers the scenario of fusing bipolar trust information from two trust networks. It selects minimum of both the values from each of the network as the potential trust value.

2.5 Semantic Trust Modelling

Trust information in consolidated semantic networks is represented using ontology and the semantic web provides a family of such knowledge representation standards. All these are W3C recommended, with Resource Description Framework Schema (RDFS) (Brickley and Guha, 2004) being the most simple, providing a basic set of vocabulary, with three variants of Ontology Web Language (OWL) (McGuinness and van Harmelen, 2004), OWL Lite, OWL DL and OWL Full providing an advanced set of vocabularies, with OWL Full being most powerful. The ontology proposed in this work is OWL Lite complaint, it uses *owl:sameAs* for co-referencing URIs from multiple networks.

2.5.1 Trust Ontologies

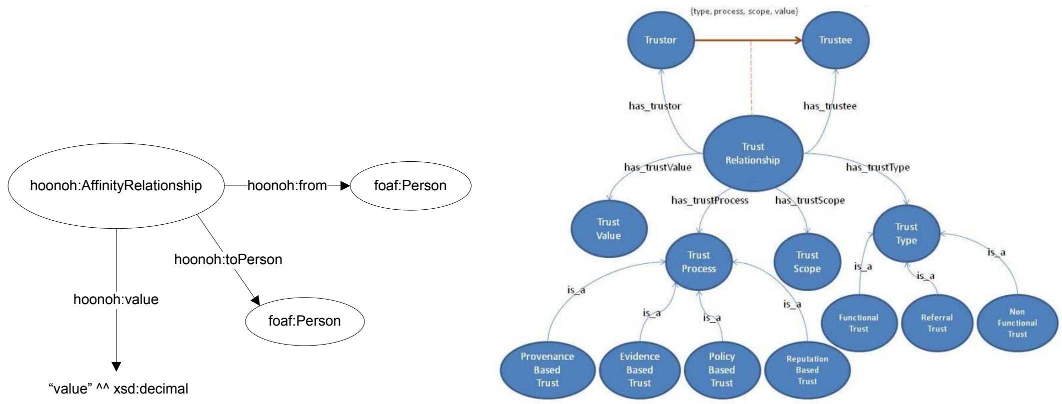
The semantic web layercake (Koivunen and Miller, 2001), presented by Berners-Lee (1998); Berners-Lee et al. (2001) and Shadbolt et al. (2006) uses ontologies for data semantics and there are number of such ontologies exist for trust. These range from interpersonal trust in social networks to managing trust and provenance information about documents on the web.

Golbeck et al. (2003) models interpersonal trust in semantic networks and describes nine different trust levels, with each level defined as a distinct property in the ontology. Although this model presents a clear description of trust, the restriction to the explicit nine levels limits its reuse in other scenarios of implementation. For example, in our case we have three fuzzy levels of trust, low, moderate and high. The ontology by Viljanen (2005) is more a generalised survey ontology, covering the various current types of trust methods, but lacking in terms of specific details about multiple network consolidation. The Web of Trust ontology deals with document-level trust and helps to annotate provenance information about documents using digital signatures[3].

The Honooh ontology (Heath and Motta, 2008) and ontology proposed by Thirunarayan and Anantharam (2011) are closer to our idea, as they model interpersonal trust and on a single network. Figure 2.5 shows that each of these ontologies describes a single *Trust* class that maps trust relationship between instances of *foaf:Person* class. Different trust properties are available that specify further details of the trust relationship. If the missing information relating to the consolidation of MuDi networks can be added, it would be in accordance with our implementation. The ontology proposed by Heath and

[3]http://xmlns.com/wot/0.1/

Motta (2008) is extended for this study by adding related properties, as described in Section 3.6.



(a) Interpersonal trust managed in Hoonoh Ontology. (b) Trust Ontology.

Figure 2.5: Two already developed trust ontology samples taken from (Heath and Motta, 2008) and (Thirunarayan and Anantharam, 2011).

2.5.2 Interlinking Multiple Networks

The consolidation of multiple RDF graphs involves interlinking resources defined in separate graphs at the instance level or schema level, and the decision on the approach depends on the semantic structure of the data (Nikolov et al., 2008; Rahm and Bernstein, 2001).

2.5.2.1 Instance Level Interlinking - Co-reference Resolution

Instance level interlinking or co-reference resolution is a technique to evaluate whether two URIs from multiple RDF graphs represent the same resource. If the URIs use the similar namespace, then it is an easy task of URI comparison, otherwise it can be carried out using two ways: **1)** logical inference or **2)** label comparison (Shi et al., 2008). Logical inference is the comparison of OWL *Inverse Functional Properties (IFPs)*[4] between resources in an RDF graphs. If matched, these properties identify URIs that represents the same resource in multiple RDF graphs.

$$\forall p / p \in IFP \Rightarrow (\forall s_1, s_2 \ / \ p(s_1) = p(s_2) \Rightarrow s_1 = s_2) \tag{2.21}$$

There are many IFPs discussed in the literature and *foaf:mbox* is an example (Hogan et al., 2007). It represents an email in the FOAF ontology, and a single email can

[4] http://www.w3.org/TR/owl-ref/#InverseFunctionalProperty-def

belong to only one person. In absence of *foaf:mbox* property due to privacy reasons, as happens in some datasets, *foaf:mbox_sha1sum* can be used and Figure 2.6 writes the rule mentioned in Equation 2.21 as SPARQL CONSTRUCT query.

```
CONSTRUCT {
  ?person1 owl:sameAs ?person2
}
WHERE {
  ?person1 rdf:type  foaf:Person .
  ?person2 rdf:type  foaf:Person .
  ?person1 foaf:mbox_sha1sum ?email .
  ?person2 foaf:mbox_sha1sum ?email .
  FILTER (?person1 != ?person2)
}
```

Figure 2.6: SPARQL CONSTRUCT query implementation of the rule in Equation 2.21, taken from Shi et al. (2008).

In situations where none of the logical comparison properties are available for co-reference resolution, label comparison can represent an alternative. It searches for string similarities between data properties (such as *foaf:name* etc) and can be categorised into three types: **1)** simple property matching, **2)** partial property matching and **3)** cross-property matching (Sleeman and Finin, 2010b). Simple property matching compares corresponding properties from two different datasets and returns 'true' if property values match exactly. Partial property matching returns 'true' if there is a partial match between string values of corresponding properties from multiple instances. The cross-property variation matches different properties from candidate datasets; for example, *foaf:nick* can be compared with *foaf:name* from another dataset. This work uses a simple property matching technique and, although it does not always give precision, it is the simplest approach and hence is used in this work as a starting point.

URI ambiguation happens if simple property matching is applied for a limited number of data properties. For example, if its only done by comparing FOAF name properties of the users. It can wrongly evaluate two URIs representing the same person, while can misclassify a pair of URIs as non-co-referred. To overcome these issues there are already developed solutions for user disambiguation. For example, Xu et al. (2015) uses contextual information other than the names for web person disambiguation. Liu et al. (2014) models it for authors of research articles and considers paper attributes in addition to the author attributes to decide about co-referred users.

Co-referred URIs can be linked using the *owl:sameAs* predicate available in the OWL Lite specification[5]. It states that the 'two URI aliases refer to the same resource'[6] and,

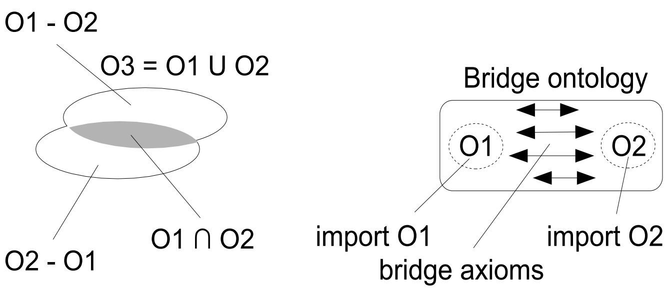
[5]http://www.w3.org/TR/2004/REC-owl-features-20040210/#s2.1
[6]http://linkeddatabook.com/editions/1.0/

once they are set as the same, they are no longer distinguishable. Due to this, in the literature there are reservations about its use (Glaser et al., 2008). Researchers believe that it is too strong to co-refer using *owl:sameAs* as it shows two URIs representing the same entity, while this is not true all the time. People have different identifications based on the context, so declaring both URIs as one loses the contextual information in which individual URIs are described. For example, there can be two URIs for Dame Wendy Hall, one as the author of a paper and another as the Head of School, and each of these URIs contain different information about, for example, emails or phone numbers. Asserting them as same using *owl:sameAs* would make it difficult to obtain specific data related to one of the URI as they are no longer differentiable. Accordingly, some researchers prefer to define their own properties (such as :coref, :Equivalent) for representation (Glaser et al., 2008; Sleeman and Finin, 2010a). However, for experimentation in this work, *owl:sameAs* is used as a starting point.

A single URI among all URIs may be marked as canonical, representing a reference that can be used for making statements about the co-referred entity. It can be selected from one of those already existing, either randomly or on the basis of occurence frequency, or it can be generated afresh using the namsespace of the resultant (named) graph (Glaser et al., 2009). This work generates a fresh URI for each co-referred participant in the consolidated named graph.

2.5.2.2 Schema Level Interlinking - Ontology Merging

Semantic merging and its related term, ontology mapping, is carried out if information from MuDi social networks uses different ontologies. It compares schema attributes (such as class name, subclass name, etc) among different ontologies and tries to find which of the classes represent the same concept or same set of entities (Kalfoglou and Schorlemmer, 2003).



(a) Ontology Merging. (b) Bridge Ontology.

Figure 2.7: Two types of ontology merging, left shows creation of new ontology O3=O1 ∪ O2 after merging, right shows creating bridging between existing ontologies O1 and O2 (De Bruijn et al., 2006).

There are two forms of ontology mapping. Either existing similar ontologies are bridged by creating links between them, or a new ontology is generated by merging existing ones. Contrary to its name, *Ontomerge* (Dou et al., 2002) bridges ontologies specifying similar constructs in existing ontologies, while *PROMPT* (Noy and Musen, 2001) takes two ontologies as an input and generates a single merged ontology by replacing all the similar classes with individual ones.

If the data uses the same vocabulary for information representation (for example, user data in FOAF format) across separate graphs, then instance-level matching is sufficient to interlink multiple networks. If the data is modelled using different ontologies, then the schema level matching needs to be done before running any instance level co-referencing algorithm (Sleeman, 2012). The current implementation of this work assumes FOAF data is available from MuDi networks and hence performs instance level co-reference for interlinking networks. To incorporate data from networks with varied ontologies, an ontology mapping layer would be established in future.

2.5.3 Trust Annotations

RDF trust statements in consolidated graphs can be made by asserting statements about existing relationships through a process known as reification. If there is no such relationship to be reified, new assertions based on the ontology can be made, and Section 3.6 proposes an ontology for aggregated trust over MuDi networks.

2.5.3.1 RDF Reification

RDF Reification (Hayes, 2004; Futrelle, 2006) is the mechanism for making statements about existing triples in the RDF graph. It starts by making an instance of *rdf:Statement* class with *subject*, *predicate* and *object* of an existing triple, becoming three distinct triples in the reified statement and increasing the minimum number of triples to four. There can be many situations where people use this technique. For example, to reify the graph for its creator, a set of triples is included, as shown in RDF Graph 1 (taken from[7]).

Figure 2.8 shows that it takes an existing triple '*a exproducts:item10245 having weight of 2.4*' and extends it for its creator with '*exstaff:85740*'. *exproducts:item10245*, *exterms:weight* and *2.4* in existing triple are added as three separate triples with *rdf:subject*, *rdf:predicate* and *rdf:object* predicates, respectively. A reified triple with *dc:creator* property is then added in addition to the four triples.

There are two approaches when specifying the subject of triples in the reified statement. It can either be a URI derived from a namespace, as in RDF Graph 1, or it can be

[7] http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#reification

specified as a blank node. Blank node is an RDF node usually with a local identifier that does not contain any data in itself, but serves as the parent node to group other data. The use of blank nodes is discouraged because absence of the global identifier (URI) limits its scope to the specific graph. If that graph is to be combined with some other graph, then the data about same resources would not be merged. There are a number of issues relating to the implementation of reification that hinders its use in trust modelling over MuDi networks and these are specifically discussed in Section 3.5.1.

RDF Graph 1 RDF Reification Example

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix exterms: <http://www.example.com/terms/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix exstaff: <http://www.example.org/staffid/>
@prefix exproducts: <http://www.example.com/2002/04/products#>

exproducts:item12345 rdf:type rdf:Statement.
exproducts:item12345 rdf:subject exproducts:item10245.
exproducts:item12345 rdf:predicate exterms:weight.
exproducts:item12345 rdf:object "2.4"^^xsd:decimal.
exproducts:item12345 dc:creator exstaff:85740 .
```

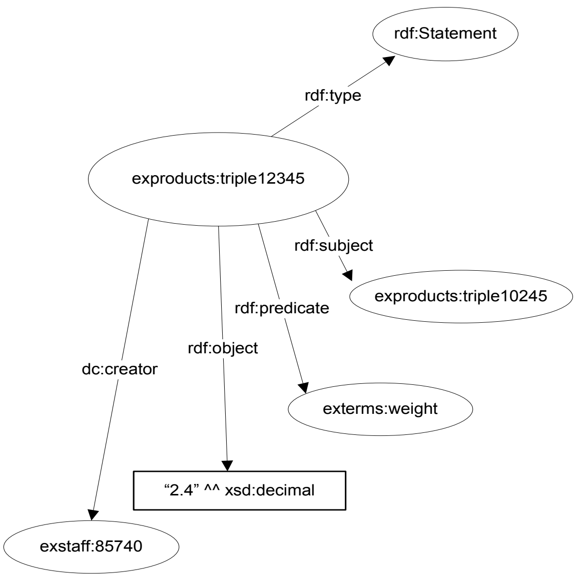


Figure 2.8: RDF Reification example taken from 7.

2.5.3.2 Adding New Assertions

Adding new assertions is another way of annotating trust information in the graph. This includes proposing new ontology, reusing URIs of the participants from an existing graph and adding trust properties between them using the proposed ontology. Unlike reification, this method is a syntactic level representation of trust data, and it can be reused and queried easily.

A number of implementations have used this technique for modelling subjective trust between users, for example, Golbeck et al. (2003) annotate trust and distrust information (shown in Figure 2.9) between users *Bob* and *Dan* by proposing an ontology that assumes *foaf:knows* relationship between them in an existing network. It shows that the *Bob highly trusts Dan regarding Research but distrusts him for Auto Repair*. Similarly, the Hoonoh model (Heath and Motta, 2008) annotates an existing FOAF graph using the ontology mentioned in Figure 2.5(a) and a sample annotated fregement is shown in Figure 2.10 for two users, *abc123* and *xyz789*.

```
<Person rdf:ID="Bob">
        <mbox rdf:resource="mailto:Bob@example.com"/>

        <trustsHighlyRe>
                <TrustsRegarding>
                        <trustsPerson rdf:resource="#Dan"/>
                        <trustsOnSubject rdf:resource="http://example.com/ont#Research"/>
                </TrustsRegarding>
        </trustsHighlyRe>

        <distrustsAbsolutelyRe>
                <TrustsRegarding>
                        <trustsPerson rdf:resource="#Dan"/>
                        <trustsOnSubject rdf:resource="http://example.com/ont#AutoRepair"/>
                </TrustsRegarding>
        </distrustsAbsolutelyRe>

</Person>
```

Figure 2.9: Adding trust annotation method used by Golbeck et al. (2003).

```
<?xml version="1.0" encoding="UTF8"?>
        <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22rdfsyntaxns#"
                xmlns:rdfs="http://www.w3.org/2000/01/rdfschema#"
                xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
                xmlns:owl="http://www.w3.org/2002/07/owl#"
                xmlns:hoonoh="http://hoonoh.com/ontology#"
                xmlns:foaf="http://xmlns.com/foaf/0.1/"
                xml:base="http://hoonoh.com/">

        <hoonoh:AffinityRelationship rdf:about="relationships/affinity/abc123/xyz789">
                <hoonoh:from rdf:resource="people/abc123"/>
                <hoonoh:toPerson rdf:resource="people/xyz789"/>
                <hoonoh:value rdf:datatype= "http://www.w3.org/2001/XMLSchema#decimal"> 0.8500 </hoonoh:value>
        </hoonoh:AffinityRelationship>

        <foaf:Person rdf:about="people/abc123">
                <foaf:mbox_sha1sum>abc123</foaf:mbox_sha1sum>
        </foaf:Person>

        <foaf:Person rdf:about="people/xyz789">
                <foaf:mbox_sha1sum>xyz789</foaf:mbox_sha1sum>
        </foaf:Person>

</rdf:RDF>
```

Figure 2.10: Trust annotations for sample users using Hoonoh ontology (Heath and Motta, 2008).

2.6 Semantic Trust Management

Trust annotations can be made in existing graphs, then republished with trust annotations, or as separate graphs having trust information with URIs of the resources linking to the data in the original graphs. Triplestores let us manage these graphs as separate storage units and allow us to query a specific named graph.

2.6.1 Named Graphs

Named Graphs (Carroll et al., 2005a) assign a URI to each RDF graph and can be seen as a semantic data management construct that allows us to maintain data from multiple sources as a separate set of graphs with URIs representing sources of the graph. It allows us to add provenance information about these graphs and lets us query information and apply procedures specific to relevant graphs.

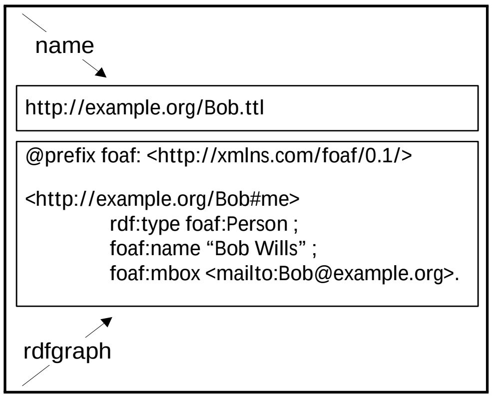


Figure 2.11: A sample named graph *ng*.

At an abstract level, a named graph can be seen as a pair *ng=(n,g)* with n representing the name of the graph in the form of a URI (formally written as *name(ng)=n*), while g shows the rdfgraph having actual information in the form of triples (formally written as *rdfgraph(ng)=g*). Figure 2.11 shows a sample named graph having FOAF file of the person *Bob* with name(ng) = *http://example.org/Bob.ttl* and rdfgraph(ng) is the set of triples in the Turtle format. The scope of the triples in such graphs is limited to that specific graph, and multiple graphs can hold same data with different names. If merged, then the data about the same resources will be fused, due to having similar URI in both the networks. However, this does not happen for blank nodes; blank nodes defined in two graphs *ng* and *ng'* remain disjoint sets (Kleyn and Carroll, 2004) due to the absence of a single assigned URI.

Named graph use for recording provenance and trust information is discussed in the literature. Carroll et al. (2005b) and Watkins and Nicole (2006) have used it for recording provenance information about graphs, thus establishing whether or not to trust the information provided by this source. This technique is also presented as an alternative of the reification as, unlike reification, it is a structural way of presenting information.

There are two serialisation formats for named graphs, XML format for serialisation, TriX, proposed by Carroll and Stickler (2004); and textual human friendly syntax, TriG (Bizer and Cyganiak, 2013). TriX provides an alternative to RDF/XML, using DTD and XML for specifying the name and triples of the named graph while TriG extends Turtle by adding { and } to group triples in multiple graphs, with each graph preceded by the name of that graph (example in Figure 2.12). Due to its plain, easily understandable format, TriG will be used in this work to describe triples in the named graphs.

2.6.1.1 Graph Per Source

One application of the named graph is to represent data from different sources as distinct named graphs with resolvable URIs (Dodds and Davis, 2011). This will work both for making statements about information source and as a URL to harvest data, if it is available as an RDF document. Figure 2.12 shows a sample RDF document from the web, it can be stored as a named graph with source URI as name(ng) = <*http://www.example.org/person.rdf*> in the triplestore.

```
#Named graph URI is source document
<http://www.example.org/person.rdf> {
#Triples from source document
        <http://www.example.org/person/joe> foaf:name "Joe Bloggs".
}
```

Figure 2.12: Graph per source, example taken from Dodds and Davis (2011).

For trust over MuDi networks, this technique creates an opportunity to publish consolidated networks with aggregated trust information as a distinct named graph with the linked data statements pointing towards individual graphs for further information (using *owl:sameAs* or *owl:seeAlso*).

2.6.1.2 Graph Per Aspect

The concept of named graphs can be used to generate separate graphs for different aspects of the information from the dataset (Dodds and Davis, 2011). It separates out information from different sources about different resources and puts them in a new named graph that satisfies that aspect. For example, in Figure 2.13, each of the three resources are described for their *foaf:primaryTopic* in a graph with name <*http://data.example.org/graphs*>.

In the consolidated MuDi trust, this technique can help in finding out those participant pairs who have trust relations between them in the area of *Research*. Similarly, those who have a high degree of trust between them across all the MuDi networks can be another information aspect.

```
#core description of a resource; provided by user
<http://data.example.org/graphs/core/document/1> {
        <http://example.org/document/1> dct:title "Bath in the Summertime".
}

#tags; maintained by process 1.
<http://data.example.org/graphs/tags/document/1> {
        <http://example.org/document/1> dc:subject "Bath".
        <http://example.org/document/1> dc:subject "Travel".
}

#related links; maintained by process 2.
<http://data.example.org/graphs/links/document/1> {
        <http://example.org/document/1> dct:related <http://travel.example.org/doc/bath>.
}

#System metadata graph, listing topic of each graph
<http://data.example.org/graphs> {
        <http://data.example.org/graphs/core/document/1> foaf:primaryTopic <http://example.org/document/1>.
        <http://data.example.org/graphs/tags/document/1> foaf:primaryTopic <http://example.org/document/1>.
        <http://data.example.org/graphs/links/document/1> foaf:primaryTopic <http://example.org/document/1>.
}
```

Figure 2.13: Graph per aspect, example taken from Dodds and Davis (2011).

2.6.2 Sesame TripleStore

Sesame (Broekstra et al., 2002, 2003) stores named graphs as *quads* with an added *context* against each triple in the repository (Cyganiak et al., 2012). *Context* specifies the background of the triple information and is extended in this work to provide the sources of multiple social networks.

$< subject >< predicate >< object >< context >$

RDF data from the MuDi networks can be added into this store with triples from the same graph having a similar context, and SPARQL queries can be run for performing structure and syntax level operations (such as instance level co-referencing, etc). A trust-annotated version of the consolidated named graphs can be stored as a separate graph with a different *context* value, and newly generated URIs point to co-referred URIs in the individual graphs.

To query information from contextual graphs, the SPARQL construct can be used with *context* of the graph mentioned using keyword *GRAPH* as shown in SPARQL SELECT Query 2. An implementation of these graphs is discussed in Section 3.5.3 in the context of trust annotated linked named graphs.

SPARQL SELECT Query 2 Sample SPARQL query for named grapphs to sesame

```
@prefix ex: <http://example.org/>.

SELECT ?person ?email
WHERE {
GRAPH ?G1 {
        ?person ex:email ?email
}
GRAPH ?G2 {
        ?G1 ex:author ex:Chris
}
```

2.7 Expert Recommendation Mechanisms

Consolidation of the MuDi networks also creates opportunites for generating trust-aware expert recommendations spanning multiple social networks, as people prefer to interact with those with a history of good interactions or recommended by other trustworthy experts. That expert can be in the same network as the querying user or in some other network, now accessible after interlinking them.

Expert recommendation systems can be classified as techniques based on merely analysing research profiles of users, and techniques of analysing both research profile information and social network information. Crowder et al. (2002a,b) describes expert recommendation models for corporate organisations where information about experts is stored at a single location or predetermined set of locations with shared attributes (such as email, publications, etc). This avoids the problem of consolidating user information from multiple networks with the varied identities of each of the network before analysing the expertise of the users. A ranked list of people is returned, based on the publications, with the person having the most publications being considered as the chief expert in the field.

Zhang et al. (2007) and Li et al. (2007) consider both profile information and links with other researchers in the field for expert recommendations. An expert score is calculated using probabilistic relevance between searched query and documents authored by researchers; those having a high value are considered to be the expert in that topic. To improve the accuracy of the system, a social network is generated using co-author relations, and if a researcher knows many such people then this increases their likelihood of being an expert in that field. Expert scores available on the links between researchers also propagate along the paths (as trust transitivity exists with decay in social networks) with a predetermined weight known as the propagation cofficient (in the range 0 to 1) (Zhang et al., 2008).

2.8 Conclusions

This chapter situates our study in the literature, and describes the relevant technologies and techniques used in subsequent chapters. Trust is defined as having a subjective value that carries a personalised local content extracted from explicit and implicit social networks. It follows the trust transitivity rule and there is decay of trust along the paths.

When evaluated over MuDi social networks, a concept of *local consolidation* of trust exists, rather than a *global consolidation*, as it results in aggregating trust information between pairs of participants available from the links between them. It can be scoped down to *subjective local consolidation* by consolidating networks related to specific fields, as has been done in this work, by analysing it for a dataset from professional social networks. So the definition by Olmedilla et al. (2006) to conceptualise it as a local metric is adapted for this work:

Definition: *Trust of party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period (within a specified context in relation to service X).*

Co-reference resolution identifies similar users across multiple networks and data fusion techniques aggregate overlapping trust information on links between them. Semantic technologies help the modelling and management of trust in machine-readable format that can be published on the web for others to consume.

The next chapter uses this idea, and establishes a semantic framework for capturing and manipulating trust networks that can be used to consolidate MuDi networks and act as the basis of MuDi trust calculations.

Chapter 3

A Semantic Web Framework for Consolidating Multiple Social Networks

3.1 Introduction

This thesis proposes to improve trust metrics by drawing on multiple rather than a single social network. To do this the networks must be consolidated, which means using a common representation and a consolidation technique. This chapter explains the challenges in consolidating MuDi networks, describes a mechanism for resolving the identities of users from different social networks, and discusses different semantic annotation techniques for annotating trust data in the consolidated MuDi graph. The proposed trust ontology provides the solution for making trust assertions by defining a class and a set of properties related to the consolidated trust. The annotated trust graph is published as a separate named graph for others to use.

3.2 MuDiTCF Architecture

Figure 3.1 presents the layered architecture of the proposed MUltiple DIstributed Trust Consolidation Framework (MuDiTCF) for building trust and expert recommendation applications over multiple distributed social networks. Semantic data available from diverse activities of users on the web is passed through sequential components of the framework to generate a consolidated network that can be used to calculate measures of interpersonal trust.

There are many systems that provide specific solutions for the functionalities described in each step of the framework, such as systems for co-referencing, data fusion and trust evaluation, but there are no systems that use these techniques to build linked data applications for trust over multiple social networks. This framework specifically targets this opportunity and uses the advantages of semantic web technology to enable trust applications drawing on heterogeneous networks.

3.2.1 Framework Description

The framework is built on top of the Sesame triplestore, and it uses a number of processing modules written in Python to manipulate RDF gathered from multiple sources, consolidates this into a single representation, and then applies trust evaluation algorithms.

The ***data acquisition module*** extracts RDF data between users from heterogeneous resources over the web using SPARQL wrapper classes. It specifically aims to find the information on which trust is based and presents a network of people connected with each other, having trust ranks on the links between them. Currently, it only considers local or remote triplestores, but it is proposed to crawl RDF files as well, in future. Section 3.6 describes the ontology we developed to model these networks.

Extracted data is analysed by the ***co-reference module*** to identify whether it includes URI aliases representing people with different identities in multiple networks. It uses an heuristic approach of label comparison and returns a new canonical URI for annotating data about the user in the consolidated networks system (Sleeman and Finin, 2010b; Glaser et al., 2009). Section 3.4 describes the co-referencing mechanism in detail.

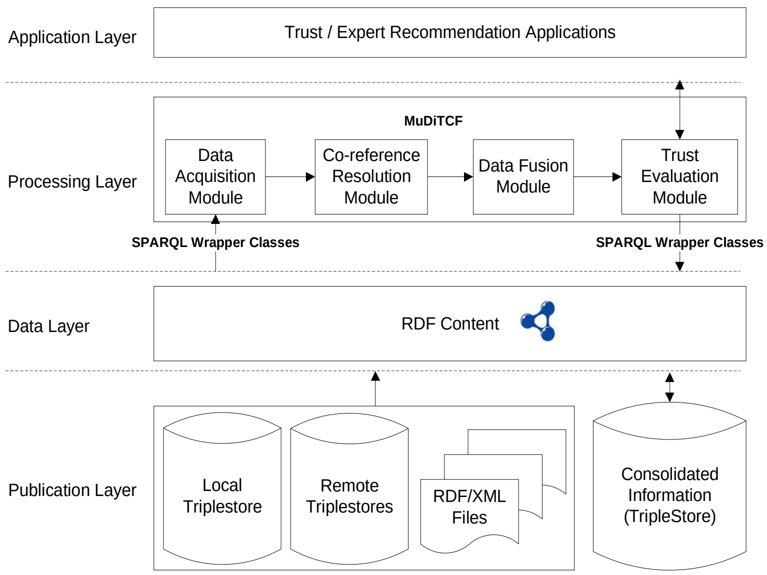


Figure 3.1: The Semantic Web Framework for building trust and expert recommendation applications over multiple distributed social networks

Aggregation of the trust information between users from multiple networks is carried out in the ***data fusion module***. It takes all available data points between a pair of users into the system, considers the importance of these values along with their sources, and generates a single value using data fusion techniques (Yager, 1988; Vicenc, 1997). These are discussed in more detail in Section 4.5.

This results in a consolidated trust network that is then sent to the ***trust evaluation module***, which can then calculate interpersonal trust values for any two individuals in the network. It prefers direct information, if available; otherwise it uses trust paths in the consolidated network to calculate the trust values. The module uses established algorithms for finding strongest and shortest trust paths (Lesani and Bagheri, 2006; Walter et al., 2008). This is explored more in Section 4.7.

The resultant trust information about users can then be annotated in the triplestore, publishing it as a named graph. This information can be re-evaluated by integrating further information from the web to refine existing trust metrics.

3.2.2 Example

The following example demonstrates how the framework is used in practice. Suppose two networks A and B (shown in Figure 3.2) needs to be consolidated, there are two types of links carrying trust information between participants represented as *t1* and *t2*. To analyse these networks for consolidated trust information, first they are searched for co-referred users, replacing multiple identifiers of users in networks with a single one. This results in people connected in multiple networks having multiple links between them, and these need to be consolidated into a single trust value. Data fusion does this by generating composite values for each such link, represented as *t1ot2* on the links in the Figure 3.2.

Consolidation of multiple networks generates trust paths between those individuals in different networks, other than opening up new trust paths between those present in the same network. For isolated users, it creates an opportunity to explore and interact with trusted people related to different areas that would not otherwise have been possible in a single network. Such a sample value exists between Users *1* and *8*, represented in Figure 3.2 as a new type of link *t3* that was not present in either of the existing networks. Those in the same network are connected through more potential trust paths than they were in individual networks. For example, in Network A, there was only one trust path ($3 \rightarrow 2 \rightarrow 4$) between Members *3* and *4*, but after consolidation the number of trust paths increases to *3* ($3 \rightarrow 2 \rightarrow 4$, $3 \rightarrow 8 \rightarrow 4$, $3 \rightarrow 9 \rightarrow 8 \rightarrow 4$).

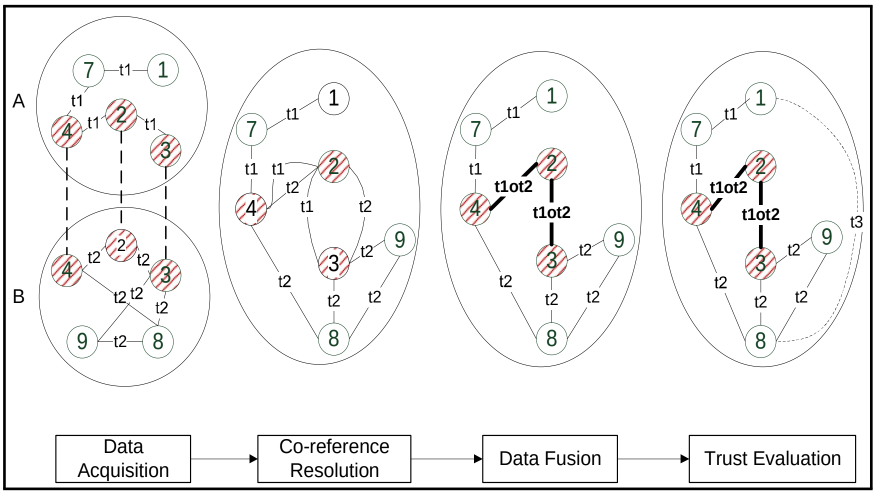


Figure 3.2: MuDiTCF example for a set of two networks (A and B) that shows the working of each component of the framework. Pattern filled nodes represent participants part of both the networks and thicked lined links represent fused information from these networks.

3.3 Challanges in Consolidating Semantic Networks

The key to the semantic web is that each resource should have a unique identifier (URI), but in social networks different networks have different namespaces and this results in people having distinct URIs in different social networks. Hence, a mechanism needs to be created that can retrieve URIs co-referring to the same person and replace those URIs with a single one for making annotations.

An ontology for making trust annotations and a mechanism to manage individual and consolidated trust networks also needs to be devised. An ontology can help us in translating trust data generated from different networks by providing a uniform format. Managing individual and consolidated versions of graphs would also help to scope down methods to a network level, rather than running queries over large datasets.

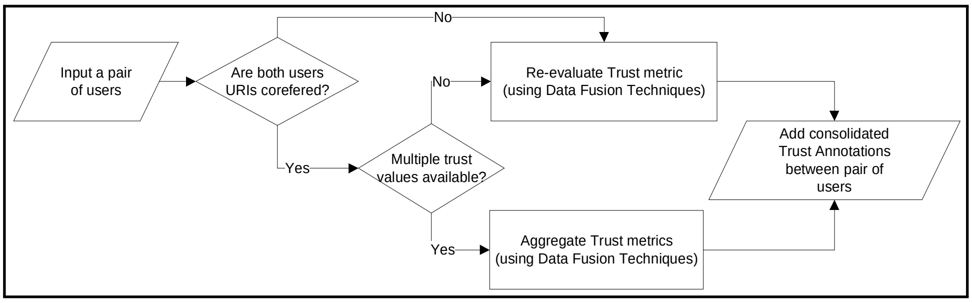


Figure 3.3: Flowchart describing steps of consolidating semantic networks

Figure 3.3 describes the key operations of the consolidation process as a flowchart. A pair of users from a social network is input into the system. The system tries to find whether these users exist in any of the other social networks, using co-reference resolution (discussed in Section 3.4). If so, then it searches for the trust information between the user pair from those networks. If such information is present, then it results in multiple trust values that need to be aggregated, otherwise the singular trust information is just re-evaluated due to the non-availability of information from other networks. If the user pair does not exist in any of the other networks, the singular trust information is just re-evaluated for a final trust value. Information aggregation and re-evaluation is conducted using various data fusion techniques. The calculated final value is then added into a trust graph as a new trust annotation using the ontology (a proposed ontology is discussed in Section 3.6).

3.4 Co-reference Resolution

When consolidating MuDi social networks, co-reference resolution is needed. This locates different URIs in multiple networks that represent the same person, and consolidates them into one new URI in the consolidated MuDi networks. Figure 3.4 shows the architecture of carrying out co-reference resolution in this work. URI pairs from the MuDi social networks are input to the classifier module of the system. It uses rule-based model and groups the candidate URIs into two categories, co-referred and non-co-referred. The co-referred category comprises users who are part of MuDi networks and they represent overlapping users, while the non-co-referred category contains users who are only present in one of the social networks. The classifier uses the rule mentioned in Equation 3.1 and compares owl Datatype Properties (i.e. *owl:DP*) between users present in multiple social networks. If there is a common value of data property, $?x$, between users $?a$ and $?b$ in different networks, then $?a$ and $?b$ are inferred to be co-referred users.

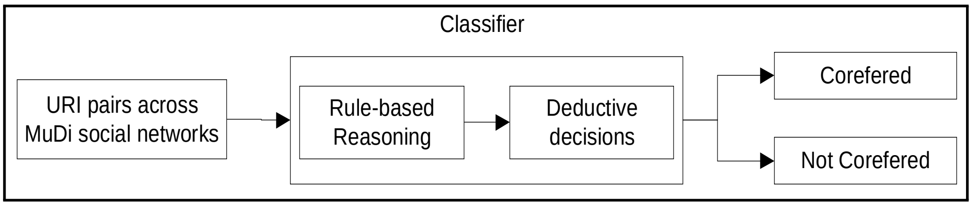


Figure 3.4: Co-reference Resolution Architecture

$$\{?p \quad a \quad owl:DP. \quad ?a \quad ?p \quad ?x. \quad ?b \quad ?p \quad ?x.\} \Rightarrow \{?a \quad :corefer \quad ?b.\} \tag{3.1}$$

Applying it in the context of trust over MuDi networks, there exists name ambiguation problem when data properties are compared. Same users can have name variants in different social networks which can result in wrongly classifying them as non-co-referred users. Similarly, different individuals can have same name and data property comparison can result in wrongly classifying them as co-referred users. So the problem is twofold as these two issues can severely affect precision of the system. The first problem can be resolved by already defined name disambiguation algorithms designed for author disambigaution in research articles. There are many such systems discussed in Section 2.5.2.1. Second issue can be solved by introducing more data properties for co-reference resolution. This will ensure comparatively better precision as only those having multiple matching data properties will be classified as co-referred users. There are many such implementations discussed in Section 2.5.2.1.

Co-referencing also needs to ensure both linking of consolidated data with already published individual social graphs on the web and provision for making new annotations about co-referred URIs. New annotations can be added if multiple URIs from individual graphs are replaced with a single new URI in the consolidated version of these graphs and *owl:sameAs* property can be used to link this URI with the existing ones in individual networks. Equation 3.2 shows a rule about how it can be done for two URIs *?a* and *?b* in individual networks and a newly allocated URI *?c* for them in the consolidated network. It can be seen that if both *?a* and *?b* have same *owl:DP* value *?x* in both the individual networks, then the *?c* in the consolidated version of these networks is linked with both *?a* and *?b* using *owl:sameAs* predicate, thereby stating that both these URIs are same.

$$\{?p \quad a \quad owl:DP. \quad ?a \quad ?p \quad ?x. \quad ?b \quad ?p \quad ?x.\} \Rightarrow \begin{cases} ?c \quad owl:sameAs \quad ?a. \\ ?c \quad owl:sameAs \quad ?b. \end{cases} \tag{3.2}$$

Figure 3.5 implements co-reference resolution in consolidated networks for four different URIs of a sample user, *Bob* from four sample individual networks. It can be seen that the system creates a single new URI of a type *foaf:Person* by taking hash of all the four URIs in individual networks. This new URI represents *Bob* in the consolidated version of these networks, with the trust annotations in it are made using this URI.

In this work, a heuristic technique of meta-data comparison, mentioned in Equation 3.2, is conducted using a SPARQL FILTER statement between corresponding *foaf:familyName* and *foaf:givenName* data properties, assuming that data is already resolved for name disambiguation problem. If the result of this comparison is 'true', the system assumes that both the URIs fall in the co-referred category, otherwise it goes to the non-co-referred category. The RDF named graphs for co-referred URIs mentioned in Figure 3.5 are shown both for individual and consolidated version of these networks in a TriG Implementation 3. This implementation only includes two URIs *mudig1:1124* and *mudig2:Bob* out of the four mentioned in the Figure 3.5 and are added in named graphs $< http : //mudig1.org/users/ >$ and $< http : //mudig2.co.uk/members/ >$ respectively. Co-referred version of these URIs generated by taking hash (i.e. *mudig1g2:4175e42dd975f...*) is shown in a named graph $< http : //consolidatedmudinetworks.com/mudig1mudig2/ >$. This graph also carries *owl:sameAs* statements for links with the URIs in existing networks.

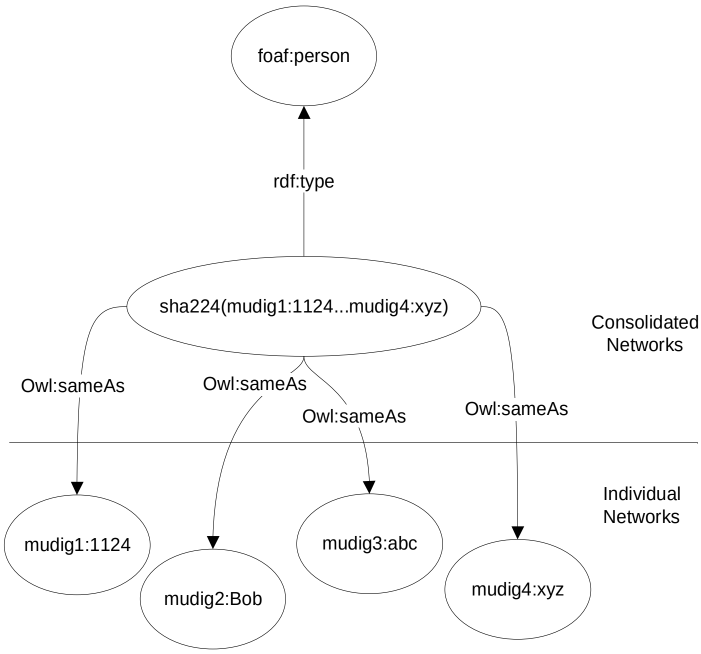


Figure 3.5: Co-reference resolution of four sample user (*Bob*) URIs belonging to four sample individual networks to generate a single URI for consolidated network. RDF statements for two leftmost URIs are shown in TriG Implementation 3

.

TriG Implementation 3 *owl:sameAs* statements for co-refered user added in consolidated named graph

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix owl: <http://www.w3.org/2002/07/owl>.
@prefix mudig1: <http://mudig1.org/users/id#>.
@prefix mudig2: <http://mudig2.co.uk/members/id#>.
@prefix mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.

<http://mudig1.org/users/> {
        mudig1:1124 rdf:type foaf:Person.
        mudig1:1124 foaf:name "Bob Wills".
        mudig1:1124 foaf:familyName "Wills".
        mudig1:1124 foaf:givenName "Bob". }
<http://mudig2.co.uk/members/> {
        mudig2:Bob rdf:type foaf:Person.
        mudig2:Bob foaf:name "Bob Wills".
        mudig2:Bob foaf:familyName "Wills".
        mudig2:Bob foaf:givenName "Bob". }
<http://consolidatedmudinetworks.com/mudig1mudig2/> {
        mudig1g2:4175e42dd975f... rdf:type foaf:Person.
        mudig1g2:4175e42dd975f... owl:sameAs mudig1:1124.
        mudig1g2:4175e42dd975f... owl:sameAs mudig2:Bob. }
```

3.5 Semantic Trust Annotations

To annotate an RDF graph with trust data, assertions need to be made about each pair of participants in the network. This can be done either by reifying existing triples (enabling the system to make statements about existing statements) or by adding new trust annotations in the form of independent named graphs.

3.5.1 Reified Trust Statements

RDF Reification is done using an instance of *rdf:statement* class. The *subject*, *predicate* and *object* of existing triple are represented using *rdf:subject*, *rdf:predicate* and *rdf:object* properties in the reified statement. Trust information can be added as a set of additional properties or relationships of the statement. For example, given two individuals, *Bob* and *Charlie*, with a *foaf:knows* relationship between them, a reified statement and *trust:absoluteValue* can be written as triples mentioned in RDF Graph 4. It can be seen that the trust value between *Bob* (having a URI *mudig1:1124* taken from one of the individual network in Figure 3.4) and *Charlie* with URI *mudig1:Charlie* is *4*. A blank node *_:aaa* is created that serves as the subject of the reified triple.

A number of issues arise when trust is managed in such a way.

RDF Graph 4 Reified Trust Statements

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix trust: <http://trustgraph.com/id#>.
@prefix mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

_:aaa rdf:type rdf:Statement.
_:aaa rdf:subject mudig1:1124.
_:aaa rdf:predicate foaf:knows.
_:aaa rdf:object mudig1:Charlie.
_:aaa trust:has_absoluteValue "4".
...
```

1. Reified statements are merely instances of built-in *rdf:statement* class. It generates triples that are just an extension of existing triples, without having any syntactical representation, as would be the case if they are added by defining a new class and property. It distorts the aesthetic of RDF and refied statements stand disconnected from the original triple. If one of them is modified, then the change is not reflected in the other triple.

2. Reification takes at least four additional triples when specifying only one trust property in the graph, and this increases the size of the dataset by at least four times. This is another drawback, because retrieving trust information from such a dataset needs to write lengthy query patterns.

3. If blank nodes are used for the subject of reified triples, then the assignment of local identifiers (rather than a URI) limits their scope to the respective graphs in which they are defined. Combining such graphs does not end up merging information about same resources, as they are considered to be different entities.

4. Reification does not allow MuDi trust networks to be published as a separate graph as these statements are added in the existing set of triples (named graphs, discussed in Section 3.5.3 allows us to do this).

5. If reification is used for trust over MuDi social networks, then the condition of having existing explicit connections between users in the network is not satisfied all the time. For example, bipartite networks (e.g. publication networks) have implicit connections between researchers, making reification incompatible with such scenarios. The dataset used for this work also extracts implicit co-authorship frequency as a metric of trust and hence there is no explicit link that can be reified.

3.5.2 Creating Semantic Silos

If existing network data and the new trust assertions are placed into a new namespace totally distinct from the published world, this creates a 'semantic silo'. It occurs when proposing a new trust ontology that not only translates existing data but gives provision for new annotations. Although this formation of data is fully functional and intelligent in its own domain, it is not connected with any other published data, which is against the spirit of linked data and the semantic web.

Implementing this idea by adding the same trust property *trust:absoluteValue* which was mentioned in Section 3.5.1, data from the existing graphs also needs to be translated into a proposed namespace (such as *mudig1g2* in this case), besides adding trust annotations.

The transformed instances of the consolidated graph and trust statement having absolute value between *Bob* and *Charlie* is shown in RDF Graph 5.

By comparing RDF Graph 5 with TriG Implementation 3, it can be seen that the *Bob* URI *mudig1g2:4175e42dd975f*, which was an instance of *foaf:Person* class, is now translated as an instance of a new *mudig1g2:Participant* class. Similary, *foaf:name* property is now replaced with *mudig1g2:name* property, with the *name* prepended with a new proposed namespace. Trust information between users is added by defining a new *trust:TrustRelationship* class and its property *trust:has absoluteValue* shows trust to be *4*.

RDF Graph 5 Trust representation in consolidated MuDi networks

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix trust: <http://trustgraph.com/id#>.
@prefix mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.

mudig1g2:4175e42dd975f... rdf:type mudig1g2:Participant.
mudig1g2:23165d34fg33c... rdf:type mudig1g2:Participant.
mudig1g2:4175e42dd975f... mudig1g2:name “Bob”.
mudig1g2:23165d34fg33c... mudig1g2:name “Charlie”.
trust:rel rdf:type trust:TrustRelationship.
trust:rel trust:has_trustor mudig1g2:4175e42dd975f...
trust:rel trust:has_trustee mudig1g2:23165d34fg33c...
trust:rel trust:has_absoluteValue “4”.
...
```

This implementation contains trust information between participants, along with other properties, in a single namespace but, if published in the current format, then it is merely an information repository disconnected from the rest of the published world. If someone has to reuse this data, schema-level matching (ontology matching) needs to be undertaken prior to the instance-level matching (co-reference resolution), as the ontology is not using any of the already developed ontologies.

Figure 3.6 shows what anyone using the newly published data by our system has to do. Ontology matching compares the classes (*foaf:Person*, *mudig1g2:Participant*) and attributes (*foaf:name*, *mudig1g3:name*) between already published individual networks and that newly published by our system. Only then can co-reference resolution find users that are part of multiple social networks. Hence, annotating data using independent trust graphs makes the reuse of data difficult and is not a suitable approach for making trust annotations.

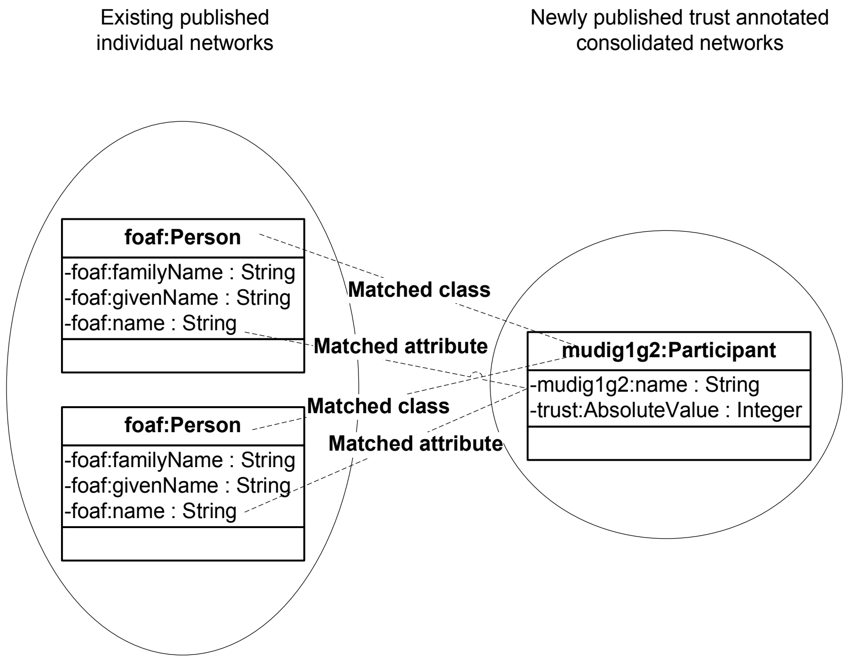


Figure 3.6: Schema matching linking consolidated trust data with existing implementations due to semantic silo

3.5.3 Linked Trust Graph

Named graphs in MuDi social networks can represent existing individual networks, and a consolidated version of those networks, as distinct graphs. For example, consider a set of named graphs, ng, having three pairs; the first pair has publication information while the second pair holds collaboration information between researchers. The consolidated version of these two graphs is a separate pair only constituting trust links between participants in both the networks. Set ng for these named graphs can be written as:

$$\begin{aligned} ng = \{ & (< http://mudig1.org/users/ >, ngmudig1), \\ & (< http://mudig2.soton.ac.uk/members/ >, ngmudig2), \\ & (< http://consolidatedmudinetworks.com/mudig1mudig2/ >, ngcmudi)\} \end{aligned} \tag{3.3}$$

where the first element of each pair shows the distinct URI of the graph for reference and the second element shows the actual RDF graph this URI is associated with. TriG

Implementation 8 (added at the end of this Section) shows serialization of these graphs as a single document, with each graph preceded by its name. It is an extension of TriG Implementation 3 and trust annotations are added between the already co-referred user, *Bob* and another user, *Charlie*. In this composition, although these graphs can have overlapping sets of participants and properties, due to each network having different URI references, they are seen as different sets of entities.

For modelling trust over MuDi networks, this named graph approach has advantages as consolidated version of trust can be seen as a separate layer of information built over existing data (as shown in Figure 3.7). Once published, it eliminates the need to write long query patterns over individual networks for co-referencing or retrieving trust information between users. Also, this layer is linked with existing published information using *owl:sameAs* statements specifying which of the URIs in the consolidated networks correspond to URIs in each of the individual networks.

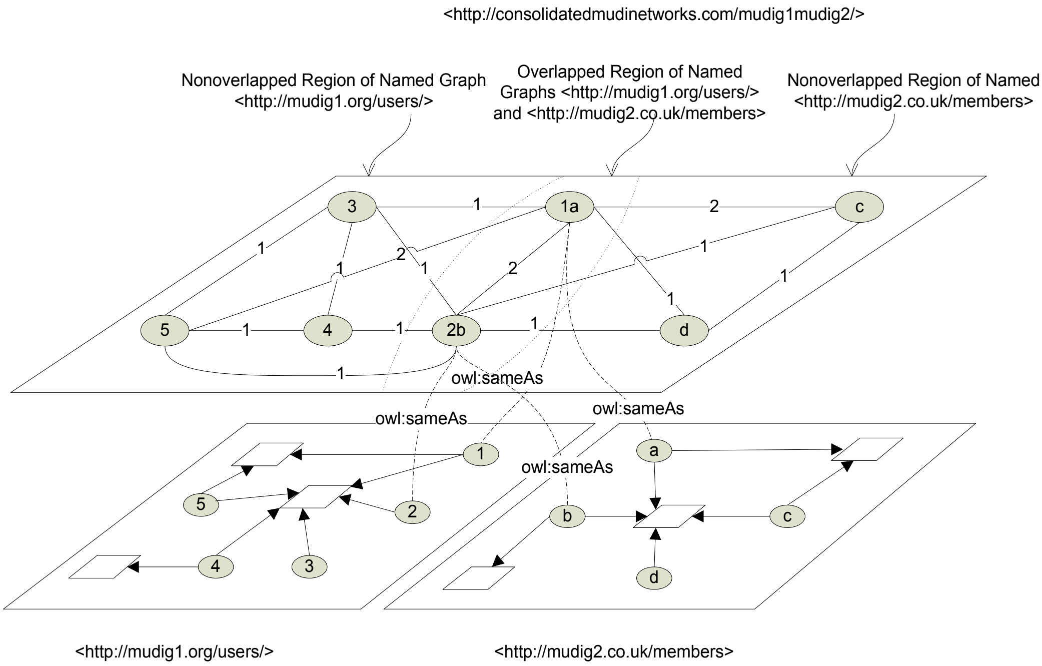


Figure 3.7: Consolidated trust network represented as an overlay network (Named Graph <http://consolidatedmudinetworks.com/mudig1mudig2/>) over individual networks (named graphs <http://mudig1.org/users/> & <http:// mudig2.co.uk/members>). *owl:sameAs* used to corefer newly assigned URI in overlapped region of consolidated graph to individual graphs.

It is also simple to annotate and query named graphs using SPARQL queries as they provide predefined constructs to deal with such graphs, as explained in Section 2.6.2. To add trust annotations in consolidated version of graphs *ngmudig1* and *ngmudig2*, one can use GRAPH statement of the SPARQL INSERT query. For example, SPARQL INSERT Query 7 adds *trust:has_absoluteValue* annotation between sample users *Bob* and *Charlie* in the consolidated graph $< http://consolidatedmudinetworks.com/mudig1mudig2/ >$ and TriG Implementation 8 shows the resultant consolidated graph.

SPARQL SELECT Query 6 Construct for retrieving trust information from named graphs

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
PREFIX trust: <trustgraph.com/id#>
PREFIX mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.

SELECT ?trustValue
FROM NAMED <http://consolidatedmudinetworks.com/mudig1mudig2/>.
WHERE {
GRAPH <http://consolidatedmudinetworks.com/mudig1mudig2/>. {
    trust:rel rdf:type trust:TrustRelationship.
    trust:rel trust:has_trustor mudig1g2:4175e42dd975f...
    trust:rel trust:has_trustee mudig1g2:23165d34fg33c...
    trust:rel trust:has_absolutevValue ?trustValue.
}
```

Similarly information can be retrieved from the relevant graph using the name of the graph in the FROM NAMED and GRAPH clause of SPARQL SELECT query. SPARQL SELECT Query 6 returns the *?trustValue* (which is added using the SPARQL INSERT Query 7) between already defined sample users *Bob* and *Charlie*.

SPARQL INSERT Query 7 Construct for asserting trust annotations in consolidated named graphs

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
PREFIX trust: <trustgraph.com/id#>
PREFIX mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.

INSERT DATA
GRAPH <http://consolidatedmudinetworks.com/mudig1mudig2/>. {
    trust:rel rdf:type trust:TrustRelationship.
    trust:rel trust:has_trustor mudig1g2:4175e42dd975f...
    trust:rel trust:has_trustee mudig1g2:23165d34fg33c...
    trust:rel trust:has_absoluteValue "4".
}
```

Named graphs can also be used to model multiple aspects of trust and this can be done by adding more named graphs into the set *ng* of the three existing ones (see Section 2.6.1.2). Existing graphs are named for their relations to different sources of information, but from within these named graphs, different ones can be generated for publishing more specific information. For example, a list of trusted people related to different research areas (such as, semanticweb, agent, etc) can be published as a more explicit dataset that allows for specific queries. Similarly, highly trusted professionals can be classified as having a separate named graph that includes people from multiple networks. Hence, the above set of named graphs *ng* can be expanded to include more graphs, keeping in view different aspects of trust:

$$\begin{aligned} ng = \{ & (< http://mudig1.org/users/>, ngmudig1), \\ & (< http://mudig2.soton.ac.uk/members/>, ngmudig2), \\ & (< http://consolidatedmudinetworks.com/mudig1mudig2/>, ngcmudi) \\ & (< http://consolidatedmudinetworks.com/semanticwebtrustedpeople>, ngrtp)\} \end{aligned} \tag{3.4}$$

This study uses the linked trust graph approach to model trust in consolidated MuDi networks. To add trust assertions into the graph, the next section proposes the consolidated trust ontology that defines trust in the context of multiple social networks.

TriG Implementation 8 Construct for adding trust annotations to co-refered users in consolidated network represented as named graph. This document is an extension of the one presented in Section 3.4 for co-reference resolution.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix owl: <http://www.w3.org/2002/07/owl>.
@prefix mudig1: <http://mudig1.org/users/id#>.
@prefix mudig2: <http://mudig2.co.uk/members/id#>.
@prefix mudig1g2: <http://consolidatedmudinetworks.com/mudig1mudig2/id#>.
@prefix trust: <http://trustgraph.com/id#>.

<http://mudig1.org/users/>{
        mudig1:1124 rdf:type foaf:Person.
        mudig1:2224 rdf:type foaf:Person.
        mudig1:1124 foaf:name "Bob".
        mudig1:2224 foaf:name "Charlie".
}
<http://mudig2.co.uk/members/>{
        mudig2:Bob rdf:type foaf:Person.
        mudig2:Charlie rdf:type foaf:Person.
        mudig2:Bob foaf:name "Bob".
        mudig2:Charlie foaf:name "Charlie".
}
<http://consolidatedmudinetworks.com/mudig1mudig2/>{
        mudig1g2:4175e42dd975f... rdf:type foaf:Person.
        mudig1g2:23165d34fg33c... rdf:type foaf:Person.
        mudig1g2:4175e42dd975f... owl:sameAs mudig1:1124.
        mudig1g2:4175e42dd975f... owl:sameAs mudig2:Bob.
        mudig1g2:23165d34fg33c... owl:sameAs mudig1:2224.
        mudig1g2:23165d34fg33c... owl:sameAs mudig2:Charlie.
        trust:rel rdf:type trust:TrustRelationship.
        trust:rel trust:has_trustor mudig1g2:4175e42dd975f...
        trust:rel trust:has_trustee mudig1g2:23165d34fg33c...
        trust:rel trust:has_absoluteValue "4".
}
```

3.6 Consolidated Trust Ontology

The trust ontology developed in this work models interpersonal trust between people over multiple social networks. It is an extended version of the ontology proposed by Heath and Motta (2008) and adds properties that are specifically needed for trust over multiple social networks. This is then coupled with the FOAF format used for the existing RDF datasets.

3.6.1 Ontology Description

The proposed OWL Lite ontology creates a single new class *TrustRelationship* that maps trust relationships between people in the network. Related to this class, two properties *has_trustor* and *has_trustee* show instances of *foaf:Person* class, with ***Trustor*** representing the trusting authority and ***Trustee*** being the one being trusted by the ***Trustor***. Here, the properties of trust are assumed to be asymmetric, meaning that trust relationship does not exist in reverse, unless explicitly stated.

There is also a set of properties (represented using *rdf:Property* in OWL Lite) attached to each relationship. The ontology allows trust values to be recorded in three different ways, namely using ***absolute***, ***processed*** and ***fuzzy*** values.

Absolute value is the most raw and the one obtained by applying the trust computation algorithm to any two users. For example, considering co-authorship frequency, it is an integer value that represents the count of number of times that a pair of researchers has appeared in publications together.

There are also scenarios when trust in the network needs to be normalised. For example, again using the example of the co-authorship frequency graph, if the highest trust value in the network is to be taken as a metric of scale (e.g. 100), and all the other metrics in the network are divided with this one, then trust between those having lowest value (e.g. 1) would almost vanish (i.e. 0.01). These estimates need to be normalised before scaling to keep this vanishing effect low. The ***processed*** value, in our model, represents such normalised values. For example, in our experimental work this normalised value is generated by adding one to the base value and taking logarithm of 10 (we add one, because the logarithm of one is zero).

Fuzzy value maps the different numerical trust values into more humanly understandable format such as, high trust, low trust and so on.

In this work, trust is defined as one of two types, ***direct*** and ***indirect*** and in this ontology it is represented using *has_type* property. Direct trust is based on direct interactions and collaborations that took place between users in the past, whereas indirect trust is calculated where direct trust values are not available. There can be different ways of doing this, for example reputation-based trust, decay-based trust and so on, but the one used in this study is the decay-based trust propagation mechanism (represented using *has_process* property). The discussion in Section 2.2.3.3 shows that propagated trust is always a weakened form of trust. For indirect evaluations, the length of trust path also matters and it is recorded using *has_pathLength* in the ontology. As trust is considered to be a subjective value, it is mentioned using *has_scope* property in the ontology. It describes the area over which this trust holds and it matters when trusted people only related to certain areas need to be searched, such as if someone is interested in academically trusted researchers in the area of semantic web or mobile learning.

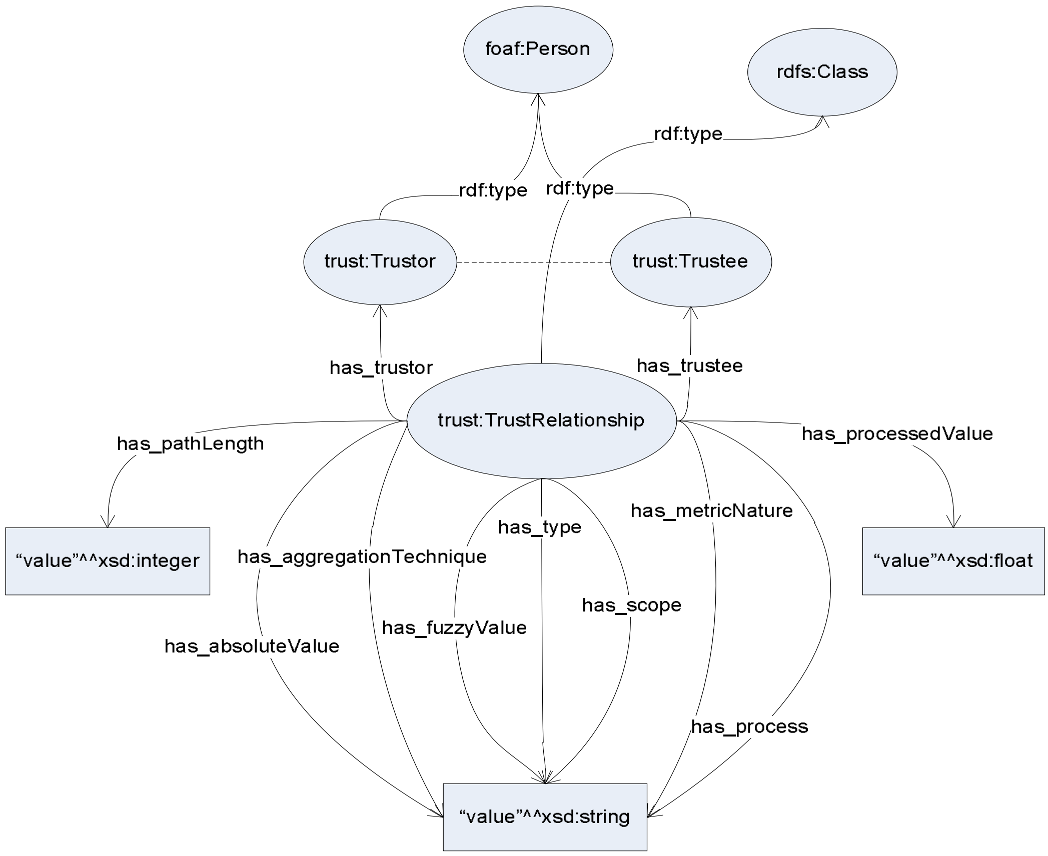


Figure 3.8: The proposed OWL Lite Trust Ontology for consolidated trust representation from MUltiple DIstributed (MuDi) social networks

Overlapping links emerge between users when trust is evaluated over consolidated social networks, and in this ontology it is modelled using two properties; *has_aggregationTechnique* and *has_metricNature*. The former represents the ***aggregation*** technique used for fusing trust values from overlapping trust links, while the later keeps track of whether this specific metric is *overlapping* or *non-overlapping*, so as to establish whether this trust metric has multiple values to aggregate, or only a single value to be re-evaluated.

3.6.2 Example

Bob and ***Charlie*** appear in a research network, they have co-authored six papers together. This compares well with the highest example in the network. In this case we might expect the following instances and relationships. Figure 3.9 shows how it appears in the form of an RDF graph.

- Trustor: $< http://mudinetworks.com/id/\_ext - 41991\_60 >$
- Trustee: $< http://mudinetworks.com/id/\_ext - 41982\_1650 >$
- Absolute Value: *6*
- Processed Value: *0.84*
- Fuzzy Value: *High Trust*
- Trust Scope: *Research*
- Trust Type: *Direct Trust*
- Trust Process: *Propagation Based Trust*
- Trust Aggregation Technique: *Weighted Ordered Weighted Averaging*
- Trust Path Length: *1*
- Trust Metric Nature: *Overlapped Trust Metric*

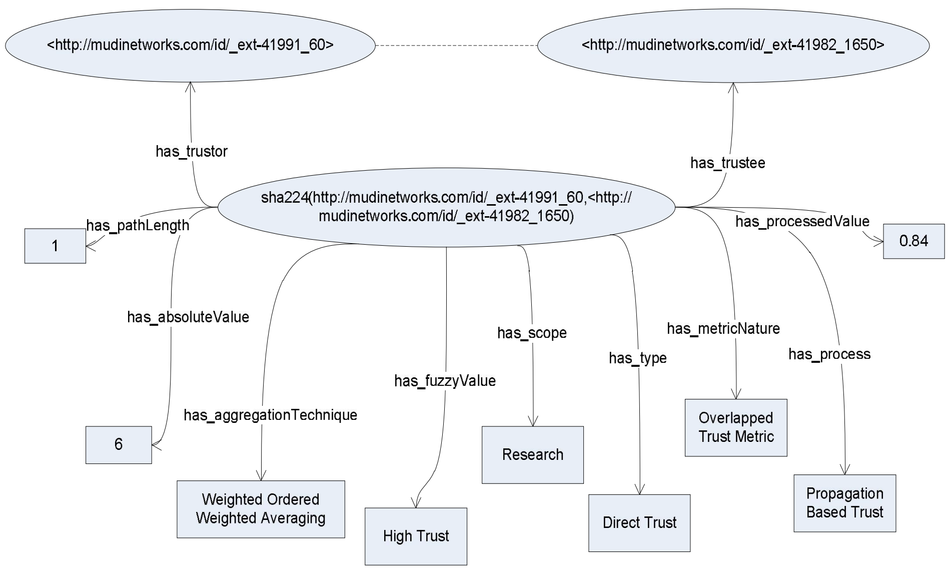


Figure 3.9: Instance diagram of the proposed trust ontology depicted in Figure 3.8

3.7 Conclusions

This chapter discusses the challenges in consolidating heterogeneous networks for trust metrics and presents a semantic web framework that provides a potential solution. The proposed framework acquires required user data, resolves identities of users from multiple networks to find co-referred participants, runs a data fusion algorithm to aggregate available trust information between them, and finally uses trust evaluation techniques for estimating trust between users. The trust ontology proposed in this work represents the consolidated trust between users in a uniform representation. It models trust as a named graph and allows the system to publish a consolidated graph as a separate information dataset with links pointing to already published information in the original graphs.

This chapter has discussed the semantic components of the MuDiTCF architecture, namely the data acquisition and co-reference resolution. The next chapter discusses mathematical algorithms relating to data fusion and trust inference over MuDi networks.

Chapter 4

Trust Data Fusion and Inference over Multiple Social Networks

4.1 Introduction

When MuDi social networks are consolidated, multiple trust values emerge between participants since they are members of multiple social networks. This chapter describes a number of trust properties that can be calculated for an aggregated version of these values. It then uses these properties to explore the impact of different (data fusion) strategies. Both naive and advanced data fusion techniques are explored and the WOWA (Weighted Ordered Weighted Averaging) technique is identified as the best for preserving trust properties. A potential trust inference mechanism for those participants not directly connected through any individual network is discussed, drawing upon existing path-finding algorithms.

4.2 Preliminaries

When aggregating multiple distributed trust between users, trust information between a pair of participants is available from n networks, resulting in trust values ranging from *one* to n that need to be consolidated. Hence, a consolidation strategy should take into account the importance of these different trust data points, along with their sources - see Section 2.4.

Let N_p represent a pair of participants and $T_{N_p} = \{T_{N_p1}, T_{N_p2} \ldots T_{N_pn}\}$ be the set of trust values between them, from n MuDi social networks. There are two sets of values associated with each data point in T_{N_p}, first is the importance of data itself, denoted as $w = \{w_1, w_2, \ldots, w_n\}$ and second is the importance of the source of that data, denoted as $p = \{p_1, p_2, \ldots p_n\}$. These values are in the range [0,1] and they represent the reliability of the trust data and the source providing that data, with values towards *one* representing higher reliability. The aggregation function, $f(T_{N_p})$, generates a single value T^{N_p} taking into account w_i and p_i for each T_{N_pi} in the data set T_{N_p}.

Table 4.1: Description of some of the preliminary set of trust parameters for multiple trust values aggregation.

Parameters	Description
$T_{N_p} = \{T_{N_p1}, T_{N_p2} \dots T_{N_pn}\}$	Set of n trust values available from n social networks
N_p	A pair of participants
$w = \{w_1, w_2, \dots w_n\}$	Set of trust data weights for n trust values
$p = \{p_1, p_2, \dots p_n\}$	Set of trust data source weights for n social networks
$f(T_{N_p})$	Function to aggregate set of trust values T_{N_p}
T^{N_p}	Aggregated trust value between N_p

4.3 Trust Aggregation Properties

When multiple trust values are aggregated to generate a single trust value, a number of trust properties can be calculated for the aggregated version of these values. These properties help to measure the impact of different aggregation functions on the resultant aggregated values in the network.

The trust aggregation function ($f(T_{N_p})$) (discussed later in the Section 4.5) generates a single consolidated value, keeping in view the values of both w and p, given that the following set of propositions should be satisfied, (adapted from (Yager, 1988; Vicenc, 1997; Yager and Filev, 1998):

Proposition 4.1 (Boundary Conditions): The trust aggregation function should be bounded; that is, the aggregated value should lie between the minimum and maximum trust data points:

$$min\{T_{N_p1}, T_{N_p2}, \dots T_{N_pn}\} \leq f(T_{N_p1}, T_{N_p2}, \dots T_{N_pn}) \leq max\{T_{N_p1}, T_{N_p2}, \dots T_{N_pn}\} \quad (4.1)$$

Proposition 4.2 (Idempotence): The trust aggregation function should be idempotent; that is, the aggregated value should be equal to T_{N_p1} if for all $x \in T_{N_p}$:

$$f(T_{N_p1}, T_{N_p1}, \dots T_{N_p1}) = T_{N_p1} \quad (4.2)$$

Proposition 4.3 (Monotonicity:) The trust aggregate function should be monotonic; that is, higher trust values should generate higher aggregated trust than lower trust values:

$$f(T_{N_p1}, T_{N_p2}, \ldots, T_{N_p1}) \geq f(T_{N_q1}, T_{N_q2}, \ldots, T_{N_qn}) \quad if \quad T_{N_pi} \geq T_{N_qi} \quad for \quad i = \{1, 2, \ldots, n\} \tag{4.3}$$

Apart from the mathematical properties, there is another property that ensures the integrity of trust from multiple social networks due to non-availability of information from certain networks. It can be interpreted as one of the three scenarios.

1. In the first case, it can be ignored as there is no information and the re-evaluation of the consolidated trust may assume a single trust value.
2. Second, it can be considered as distrust with the assumption that lack of trust has actually caused trust value to disappear.
3. In the final case and the most rational one, it can be categorised as an *absence of trust*. It emphasises that the *absence of trust* information between participants from any of the constituent social networks should not be considered as *distrust* between them.

When consolidating MuDi networks, the trust aggregation function should not consider absence of trust information from any individual network as distrust. This is due to absence of any solid evidence which should been present to distrust someone. However it should punish participants for not sharing the information. So the absence of trust appears to represent the best interpretation as it neither completely ignores the lack of data from one of the network nor it considers it distrust. It encourages people to share data which in some cases participants can deliberately avoid due to not having good reputation in that specific network. On the other hand it cannot be categorised as distrust as there is no such evidence and lack of trust information can be due to not actively using that individual network or not discovering that person in the network at all. So the decision of considering absence of data as zero is perfect choice in context of MuDi networks because still it would punish people for not sharing information but not to that extent as it happens in the case of distrust.

Proposition 4.4 (Trust Absence:) The trust aggregation function should be able to distinguish the *absence of trust* information from *distrust* between a pair of participants; that is, the aggregation of trust values with one of the values as a numeric value of zero should generate a resultant value that is approximately equal to the aggregate without that numeric zero because, for this study, numeric zero represents the *absence of trust* information not *distrust*, that is:

$$f(T_{N_p1}, 0, \ldots, T_{N_pn}) \approx f(T_{N_p1}, \ldots, T_{N_pn}) \tag{4.4}$$

4.4 Trust Aggregation Scenarios

Trust aggregation scenarios present a number of distinct types of participant pairs that emerge as a result of consolidating multiple social networks. They represent connections between participants once individual networks are consolidated as a single large graph, because after consolidation users belong to different regions of the consolidated networks. For example, in Figure 4.1, participant pairs who were just part of individual networks before consolidation now belong to regions A, AB etc and the connections between them represent new type of relationships (highlighted in different colors) between them which were non-existent in individual networks. The co-referencing resolution mechanism discussed in Section 3.4 determines such participants. Each participant pair will exist in one of the three different scenarios (see Section 1.2):

1. First type user pairs are those in the same multiple networks, i.e. both the members of the N_p exist in more than one of the consolidated networks and they exist in the same set of networks (connection tie labelled in ***red*** (in Figure 4.1(b)) represents such user pairs).
2. Second type user pairs are those who are part of one and the same individual network, i.e. each member of the N_p exists in only one of the consolidated networks, and they share the same network (connection tie between such user pairs is labelled in ***blue*** in Figure 4.1(b)).
3. Third type user pairs are those belonging to cross-regions, i.e. each member of the N_p exists in one or more of the consolidated networks, but not the same set of networks (Figure 4.1(b) depicts the connection tie between such user pairs in ***green***).

Computing the number of distinct pair combinations for each scenario in consolidated networks is non-trivial and Section 4.4.1 outlines a method of finding them.

4.4.1 Computing N_p combinations for consolidated MuDi social networks

Analysis using the venn diagram (shown in Figure 4.1(a)) reveals that, when consolidating N social networks, there are 2^N-1 type of users each belonging to one of the regions (representing different portions of the resultant consolidated MuDi networks). These users have links with other users in the same region and in other regions, and number of distinct type of user pairs for the former type can be found using Equation 4.5.

$$PC_{FRs} = 2^N - 1 \tag{4.5}$$

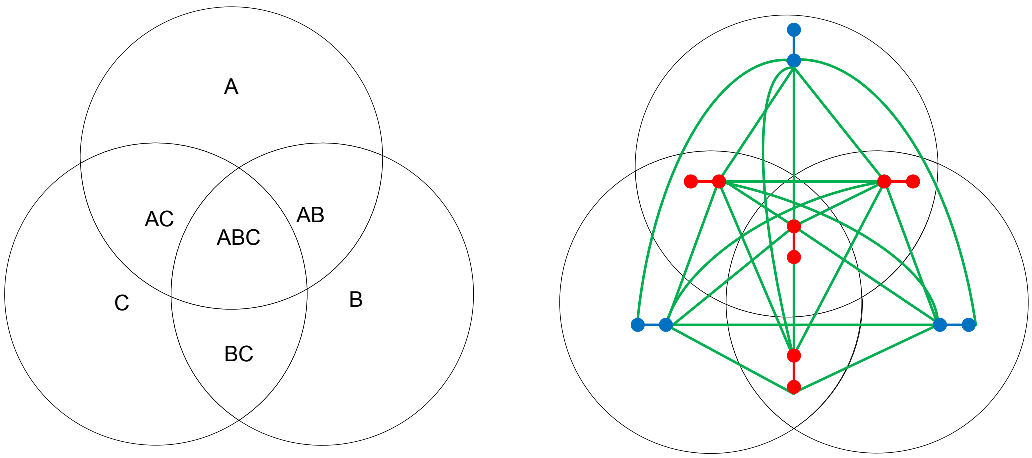
where PC_{FRs} is, *Pair Count for Fragmented Regions*, and represents the count of the distinct user pairs belonging to scenarios 1 and 2.

To analyse the same venn diagram for cross-region user pairs, distinct type of N_p combinations need to be found for the scenario 3. One of the ways is to find pair combinations, with n being the total number of regions in the consolidated graph and k being the number of members in the required combination.

$$PC_{CRs} = {}^{n}C_k = \frac{n!}{k!(n-k)!} \tag{4.6}$$

where PC_{CRs} represents, *Pair Count for Cross Regions*, and is the count of user pairs connected across different regions. The summation of values obtained from both the Equations 4.5 and 4.6 generates the total number of *Trust Aggregation Scenarios (TASs)* for the consolidated networks.

$$TASs = PC_{FRs} + PC_{CRs} \tag{4.7}$$



(a) Venn diagram representing overlapping (AB, BC, CA, ABC) and non-overlapping (A, B, C) regions of three sample social networks

(b) Detailed venn diagram with all possible pair of participants in the sample set of three networks.

Figure 4.1: Two sample venn diagrams depicting different regions and pair of participants for three sample social networks

If a similar scenario is to be implemented for asymmetric trust between users, the permutations can be used to find the possible number of cross-region participant pairs, because then the order of the users in the pair also matters as it represents the direction of trust.

$$PC_{CRs} = {}^{n}P_k = \frac{n!}{(n-k)!} \tag{4.8}$$

4.4.2 Example

Implementing the technique discussed in Section 4.4.1 for a sample set of three networks, the venn diagram in Figure 4.1(a) shows seven different regions in the resultant consolidated networks, that is, A, B, C, AB, BC, AC, ABC. The value of PC_{FRs} for these networks can be calculated by substituting N = 3 in Equation 4.5.

$$\begin{aligned} PC_{FRs} &= 2^3 - 1 \\ &= 7 \end{aligned} \tag{4.9}$$

This gives us the number of distinct type of N_P combinations that reside in any of the seven regions of the consolidated graphs with each member of the pair in the same region. Similarly, the value of PC_{CRs} for these networks can be calculated by substituting n = 7 (representing number of regions) and k = 2 (representing number of users in the N_P combination) in Equation 4.6:

$$\begin{aligned} PC_{CRs} &= {}^7C_2 = \frac{7!}{2!(7-2)!} \\ &= {}^7C_2 = 21 \end{aligned} \tag{4.10}$$

Results obtained from Equations 4.9 and 4.10 can be substituted in Equation 4.7 to calculate the total number of distinct type of N_P combinations that exists for this example of three networks.

$$\begin{aligned} TASs &= 7 + 21 \\ &= 28 \end{aligned} \tag{4.11}$$

These different *TASs* are evaluated for real-life social networks in Section 6.2.2. However, due to consolidating a pair of networks the number of scenarios for real-world networks is less as compared to the example discussed in this section.

4.5 Trust Aggregation Functions

Multiple trust values available for different *TASs* (explained in Section 4.4) need to be aggregated into a single resultant value in the consolidated version of multiple networks. This section presents different strategies for aggregating trust metrics between participant pairs available on the overlapping ties and re-evaluating those available from only one of the constituent networks. The trust metric can be classified into two different types, based on the information availability from MuDi social networks (see Figure 4.1).

1. Complete trust information available from all the networks being consolidated (region ABC).

2. Partial information from *1* to the *n-1* number of networks (for example, regions A, AB, A $\rightarrow$ AB etc).

Aggregating complete trust information (Case 1) is a straight-forward, as there are n data values, along with a similar number of sources of that information, but modalities need to be defined for aggregating partial information (Case 2). According to the trust property mentioned in Equation 4.4, there is an *absence of trust* information, so the aggregation technique should punish these partial trust metrics for the absence of information, but not to the extent that it distorts the integrity of available trust metrics. Furthermore, as mentioned in Section 4.2, there are two reliability factors associated with each of the trust data point, w and p, and the trust aggregation function should consider both these factors.

4.5.1 Weighted Averaging (WA) Aggregation

The WA aggregation mechanism aggregates trust information by considering only the reliability of the source of data. It takes the trust data set, T_{N_p}, as an input and weights each of the data point by multiplying it with the reliability factor for the corresponding source. There are two inputs to the WA aggregation function: **1)** trust vector T_{N_p}, having trust information from n social networks that needs to be aggregated and **2)** weight vector p, that contains the reliability parameter for each source.

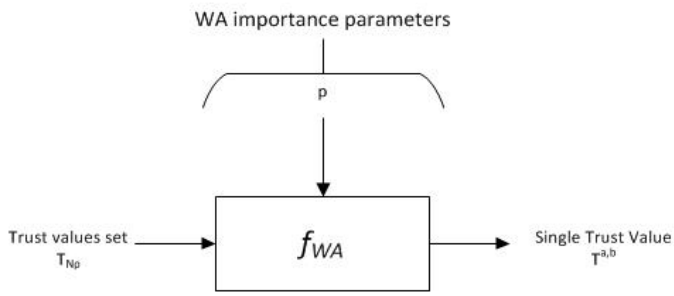


Figure 4.2: WA trust aggregation function along with input and output paramters, there is only trust data source importance parameter p.

Example:

Suppose the trust data set T_{N_p} is [0.8,0.5] and each of the sources have equal reliability, represented by the weight vector p=[0.5, 0.5]. Data in T_{N_p} can be aggregated using Equation 2.16.

$$\begin{aligned} f_{WA}(0.8, 0.5) &= \sum_i p_i T_{N_p\sigma(i)} \\ &= 0.8 * 0.5 + 0.5 * 0.5 \\ &= 0.65 \end{aligned} \tag{4.12}$$

4.5.2 Ordered Weighted Averaging (OWA) Aggregation

The OWA aggregation technique considers the importance of data as the only factor, when aggregating trust metrics from MuDi social networks. It first orders the trust data set T_{N_p} from high to low, then uses the weight vector w to weight already permuted T_{N_p} to generate a single aggregated value T^{N_p}. So there are two inputs in this aggregation function, both in the range [0,1], **1)** Trust vector T_{N_p} having n trust values and **2)** a weight vector w having n values that work as an importance factor of each value in the T_{N_p}.

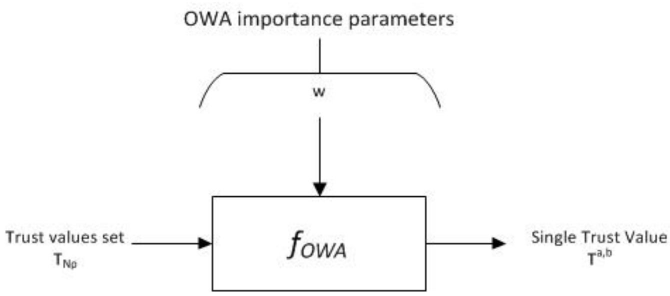


Figure 4.3: OWA trust aggregation function along with input and output paramters, there is only trust data importance parameter w.

Example:

Suppose the trust vector for this example T_{N_p} = [0.5, 0.8] and a weight vector w = [0.8, 0.2]. To aggregate this data, it first needs to be ordered as $T_{N_p\sigma}(1) \leq T_{N_p\sigma}(2) \leq \ldots \leq T_{N_p\sigma}(n)$, which results in T_{N_p} = [0.8, 0.5]. Now these ordered data values from T_{N_p} can be aggregated using Equation 2.17.

$$\begin{aligned} f_{OWA}(0.8, 0.5) &= \sum_i w_i T_{N_p\sigma(i)} \\ &= 0.8 * 0.8 + 0.2 * 0.5 \\ &= 0.74 \end{aligned} \tag{4.13}$$

4.5.3 Induced Ordered Weighted Averaging (IOWA) Aggregation

IOWA generates an aggregated trust value from multiple networks, prioritising information on the basis of the importance of data from certain networks as compared to the rest, as the data from those sources is considered to be more reliable. It takes trust data set T_{N_p} in the form of tuple $<p_i, T_{N_pi}>$ with p representing the priority order of the data with respect to the trust data source. Then it multiplies each of the ordered data point in T_{N_p} with the corresponding weight of the data provided in weight vector w.

There are three inputs to the IOWA trust aggregation function, **1)** Data set T_{N_p} comprising of trust values from n different networks, **2)** p is the order inducing vector that sets the importance order of the data points in input T_{N_p}, based on the reliability of the data source; and **3)** w specifies the weight of the data points already induced by the p. Vector p contains integer values, while the values of other two parameter are in the range [0,1].

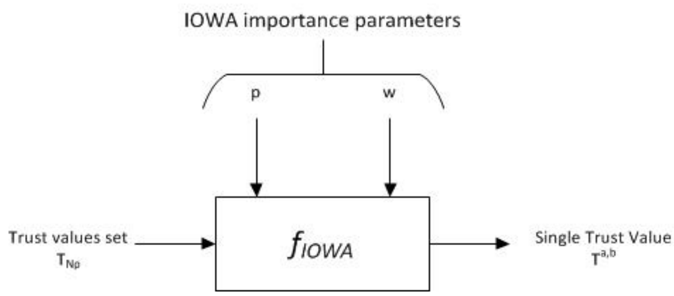


Figure 4.4: Input and output arguments of the IOWA Trust aggregation function

If, after ordering the dataset T_{N_p} based on the vector p, the resultant data order is $T_{N_p\sigma}(1) \leq T_{N_p\sigma}(2) \leq \ldots \leq T_{N_p\sigma}(n)$, then the IOWA becomes OWA. This scenario is discussed in the data example mentioned in Table 4.2.

Example:

Suppose we have the same data set as in Section 4.5.2 for aggregation, but here in the form of pairs $<p_i, T_{N_pi}>$, i.e. $(<1, 0.8>, <5, 0.5>)$ with the weighted vector w = [0.8,0.2]. Here, to aggregate this data, it first needs to be ordered in respect of p, that is, in the form of one-column matrix T_{N_p},

$$T_{N_p} = \begin{bmatrix} 0.5 \\ 0.8 \end{bmatrix} \tag{4.14}$$

Now, f_{IOWA} can be calculated using Equation 2.20,

$$\begin{aligned} f_{IOWA}(<5, 0.5>, <1, 0.8>) &= \sum_i w_i T_{N_p\sigma(i)} \\ T_{N_p\sigma} &= 0.8 * 0.5 + 0.2 * 0.8 \\ &= 0.56 \end{aligned} \tag{4.15}$$

4.5.4 Weighted Ordered Weighted Averaging (WOWA) Aggregation

The WOWA trust aggregation function also considers the reliability of both individual trust values and their sources for generating a single trust value. It takes three inputs: **1)** set of multiple trust values T_{N_p}, that needs to be aggregated; **2)** weight vector w, equal to the length of T_{N_p} that shows how reliable these set of trust values are; and **3)** weight vector p, equal to the length of T_{N_p}, to show the reliability of the sources of individual trust values. Unlike the other aggregation techniques, here w vector can have both integer and fraction values while T_{N_p} and p vectors are in the range [0,1].

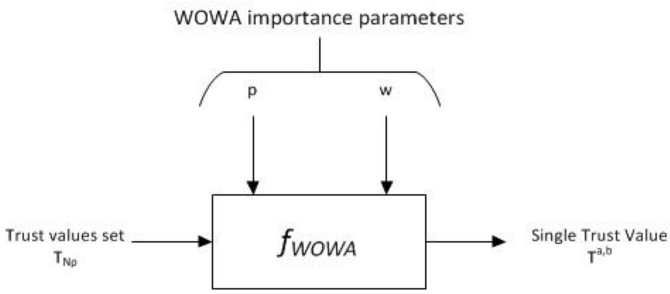


Figure 4.5: Input and output values of the WOWA Trust aggregation function

Algorithm 9 shows how weight ω_i can be calculated for each of the trust value in T_{N_p} using w and p. At *step 1* it takes T_{N_p}, w and p as inputs into the f_{WOWA} procedure. It then calculates normalised weights corresponding to each value w_i and plots these along with (0,0), i.e. $(\{(i/n, \sum_{j\leq i} w_j)|i = 1, \ldots, n\} \cup \{(0, 0)\}$ (as described in *step 2*). In *step 3*, polynomial equation w^* is derived using the curve fitting process applied on the weights plotted in *step 2*. The degree of the polynomial equation depends on the number of weight data points (plotted in *step 2*). For example if there are three data points in total, it will be second order polynomial equation. *Step 4* uses the polynomial equation w^* derived in *step 3* to find the weight vector, having weights ω_i corresponding to each of the trust values in T_{N_p}. This happens by substituting the p values into the w^* and deriving the relative importance of the trust values (a generalised procedure shown in Equation 2.19). *Step 5* orders the trust data set T_{N_p} as $T_{N_p\sigma}(1) \leq T_{N_p\sigma}(2) \leq \ldots \leq T_{N_p\sigma}(n)$. Finally *Step 6* multiplies the calculated weight values ω_i with the already ordered trust vector T_{N_p} and returns the resultant single value T^{N_p} to the calling function.

Algorithm 9 Trust Aggregation Algorithm adapted from Vicenc (1997).

```
procedure f_WOWA(T_Np: trust vector; w, p: weight vectors)
    Define S = {(i/n, Σ_{j≤i} w_j) | i = 1, ..., n} ∪ {(0,0)}
    Define w* as the function that interpolates S
    Calculate ω_i using the function w* derived in step 3.
    Order the trust vector T_Np and determine the permutation s.
    return(T^Np = Σ_i ω_i T_Np(s(i)))
end procedure
```

Example:

As an example of using WOWA, suppose that the trust information available from two MuDi networks is $T_{N_p} = [0.5, 0.8]$. The initial weights representing reliability of trust metrics are $w = [1, 0.5]$ and reliability of the sources providing that information are $p = [0.8, 0.2]$. Applying algorithm 9 to these set of data values results in following set of operations.

1. $T_{N_p} = [0.5, 0.8]$, $w = [1, 0.5]$, $p = [0.8, 0.2]$.

2. Weight vector w can be normalised as:

$$\begin{aligned} w_n &= w_i/(\Sigma_{j\leq n} w_j) \\ &= [1, 0.5]/1.5 = [0.67, 0.33] \end{aligned} \tag{4.16}$$

   Next, we need to define S by plotting the set of resultant points,

$$\begin{aligned} S &= \{(i/n, \sum_{j\leq i} w_j)|i = 1, \ldots, n\} \cup \{(0,0)\} \\ &= (1/2, w_1) = (0.5, w_1) = (0.5, 0.67) \\ &= (2/2, w_1 + w_2) = (1, w_1 + w_2) = (1, 1) \end{aligned} \tag{4.17}$$

   This gives three points {(0,0), (0.5,0.67), (1,1)} that need to be plotted and Figure 4.6 shows the plotted points.

3. There are three data points plotted in Figure 4.6 so, using the curve fitting process, the second order polynomial equation w^* derived is given below (a constant value is intentionally neglected due to its being too small).

$$w^*(x) = -0.6667x^2 + 1.6667x \tag{4.18}$$

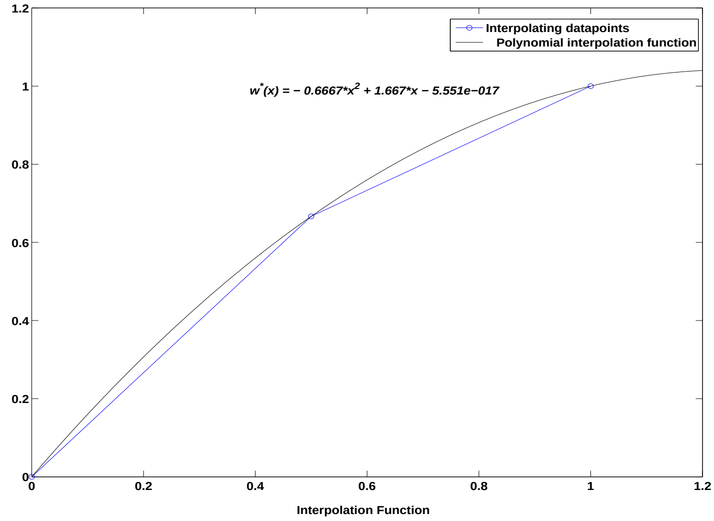


Figure 4.6: Polynomial interpolation function

4. The derived polynomial Equation 4.18 is substituted into the Equation 2.19 along with the corresponding value from the p vector. This calculates the final weights ω_i for each of the trust value in T_{N_p}.

$$\begin{aligned}\omega_1 &= w^*(\sum_{j=1}^{1} p_j) = w^*(0.8) = -0.6667 * (0.8)^2 + 1.6667 * (0.8) = 0.91 \\ \omega_2 &= w^*(\sum_{j=1}^{2} p_j) - w^*(p_1) = w^*(1) - w^*(0.8) \\ &= (-0.6667 * (1)^2 + 1.6667 * (1)) - 0.91 = 0.09\end{aligned} \tag{4.19}$$

5. Next, multiple trust metrics in trust values set T_{N_p} (from step *1*) are put in decreasing order, i.e. $T_{N_p} = [0.8, 0.5]$.

6. Finally, a single aggregated value T^{N_p} can be calculated using Equation 2.18.

$$\begin{aligned}T^{N_p} = f_{WOWA}(0.8, 0.5) &= \sum_{i=1}^{2} \omega_i T_{N_p \sigma(i)} \\ &= 0.91 * 0.8 + 0.09 * 0.5 \\ &= 0.77\end{aligned} \tag{4.20}$$

4.5.5 Comparative Analysis of Aggregation Techniques

Table 4.2 compares the aggregated values generated by WOWA and IOWA to those of the naive methods of Summation (S), Weighted Averaging (WA) and Ordered Weighted

Averaging (OWA) for a pair of trust values (it can scale up to any number of values, but here two values are shown as a sample). Data record ***2*** is the one evaluated in Sections 4.5.1, 4.5.2, 4.5.3 and 4.5.4, while the other two samples (data records ***1*** and ***3***) represent extreme conditions. Data record $[0.8, 0.8]$ shows both the values as same and high, and $[0.8, 0]$ shows one of the values as missing).

Table 4.2: Trust aggregation using five different techniques for three different set of values, 0 in $[0.8, 0]$ represents absence of trust.

No	MuDi Trust Data	S	WA	OWA	WOWA	IOWA
1	$[0.8, 0.8]$	1.6	0.8	0.8	0.8	0.8
2	$[0.8, 0.5]$	1.3	0.65	0.74	0.77	0.56
3	$[0.8, 0]$	0.8	0.4	0.64	0.72	0.16

Looking at the data with reference to our earlier assumption of respecting the integrity of high trust values, and taking zero as an *absence of trust* rather than *distrust* (the trust property mentioned in Equation 4.4), it shows that for all data records only WOWA fulfils these standards. According to the aggregation properties specified by Equations 4.1 and 4.2, the resultant value should be bounded and idempotent. S for data record ***1*** (in Table 4.2) violates both these conditions and inflates an aggregated value greater than maximum (i.e 1.6), hence becomes unfit for trust aggregation. WA values, although satisfying all the trust aggregation properties, violate trust proposition by Equation 4.4 and hence dampen down trust when a trust value is missing (such as data record ***3*** in Table 4.2, showing the aggregated value as 0.4). OWA also satisfies all the aggregation properties and considers importance of trust values, but misses the importance factor for trust data sources and hence lacks a key factor that should be considered.

WOWA removes all the shortcomings of the aforementioned techniques and considers the importance parameter for both the trust data and the trust data source. For data record ***1***, it respects the aggregation properties defined in Equations 4.1 and 4.2, and, unlike S, does not go beyond the maximum value 0.8. In the case of data record ***3***, unlike WA it senses the absence of data and punishes resultant information for not having complete data without distorting it, hence satisfies trust property specified in Equation 4.4.

Like WOWA, IOWA also considers the importance of both trust data and trust source and satisfies all the trust properties but, unlike OWA and WOWA, it allows us to prioritise data with reference to the trust source. So, it provides an additional aggregation factor missing from WOWA and can be used if trust information from a certain source is more reliable. However, if the priority of the trust source ends up permuting data $T_{N_p\sigma}(1) \leq T_{N_p\sigma}(2) \leq \ldots \leq T_{N_p\sigma}(n)$, it becomes OWA. In case, trust sources of low trust values are prioritised, IOWA can also result in dampening trust scores from individual networks. For example, data record ***3*** punishes missing data even more than WA.

Beside a set of three trust data points discussed above, there are a variety of different scenarios when consolidating two networks, such as the amount of overlap, and that in the next chapter we will examine how this impacts on the effectiveness of the various aggregation strategies.

4.6 Trust Propagation Strategies

Consolidation of multiple social networks results in creating trust paths between participants who are members of different social networks and generates new trust paths between those in the same networks. Our belief is that these trust paths will not only be shorter in length but will generate high quality trust metrics. This section discusses the two possible trust propagation strategies that will impact on the creation of trust paths between participants in the consolidated version of individual social networks.

4.6.1 Propagate Consolidate Propagate

The *Propagate Consolidate Propagate (PCP)* strategy of consolidation adopts the approach of calculating the trust between pairs of participants in their individual social networks before consolidating them into a single large social graph. This results in completely connected individual graphs, with trust information available between each pair, before consolidating them into a single network. In this case, the trust aggregation functions described in Section 4.5 aggregate the measurements between all overlapping user pairs from MuDi networks as they are presented for consolidation. After consolidation, only trust for cross-network user pairs is left for evaluation.

The networks, A and B, shown in Figure 4.7, describe the *PCP* strategy of trust propagation using a four-step procedure.

1. Firstly, individual social networks A and B acquire labelled connections between user pairs representing the direct interaction history. The connections between pairs from the Network A are labelled as *t1*, and from the network B as *t2*.

2. Then, the trust that is not available due to the absence of direct interaction between user pairs in the individual networks A and B is evaluated by applying a propagation algorithm.

3. In the third step, A and B are consolidated into a single network, with overlapping connections labelled as *t1ot2*, showing composite values from both the networks.

4. Finally, propagated trust is evaluated for isolated users and is represented as dotted links labelled *t3* label, for example, connection between users $1 \rightarrow 8$.

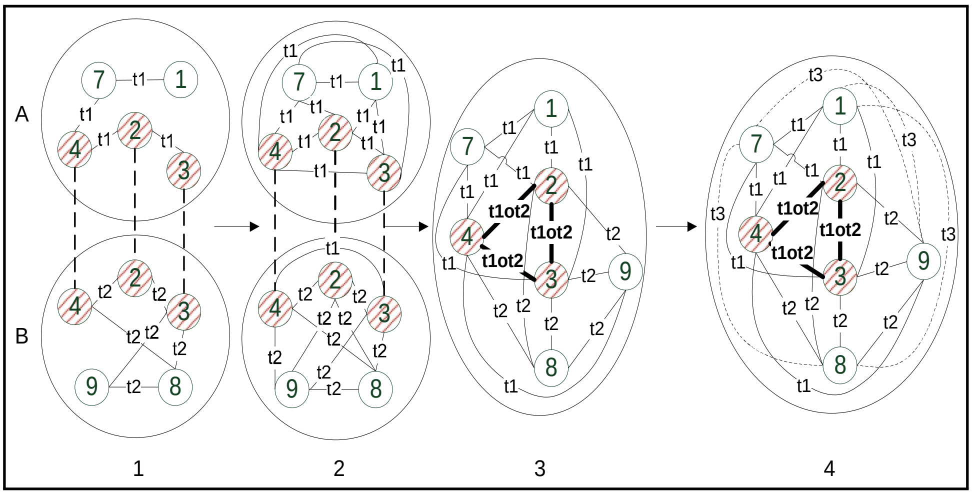


Figure 4.7: *PCP* technique that evalutes trust between indirectly connected users in individual networks. Pattern filled nodes represent users that are part of multiple social networks and bold ties represent aggregated information from both the networks consolidated.

4.6.2 Consolidate Propagate

The *Consolidate propagate (CP)* technique consolidates individual social networks into a single global network before evaluating trust for indirectly connected pairs of participants (see Figure 4.8). In this scenario, unlike the *PCP* case, indirect trust calculations are not carried out in individual networks, with the outcome that the consolidation has better and more realistic trust paths that can generate high quality trust values.

The same networks *A* and *B*, as used in the *PCP* strategy, are reused here for the description of the *CP* trust propagation strategy. Figure 4.8 describes the working of this strategy using three steps.

1. Firstly, individual social networks *A* and *B* acquire labelled connections between user pairs representing direct interaction history. The connections between pairs from Network *A* are labelled as *t1* and from Network B as *t2*.

2. Then networks *A* and *B* are consolidated into a single social graph, with connections from both the individual networks and the connections having composite value being represented as *t1ot2*.

3. In the final step, trust is evaluated between isolated pairs of participants in the consolidated graph, represented as *t3* type links in Figure 4.8.

In the CP technique, unlike the PCP technique, there is a possibility of many new trust paths emerging between participants that were not there in individual networks before consolidation. This strategy will not only impact on the accuracy of trust metrics but will help in exploring the true potential of consolidating these MuDi networks. For this reason, for this work we adopt the CP strategy of consolidating multiple social networks.

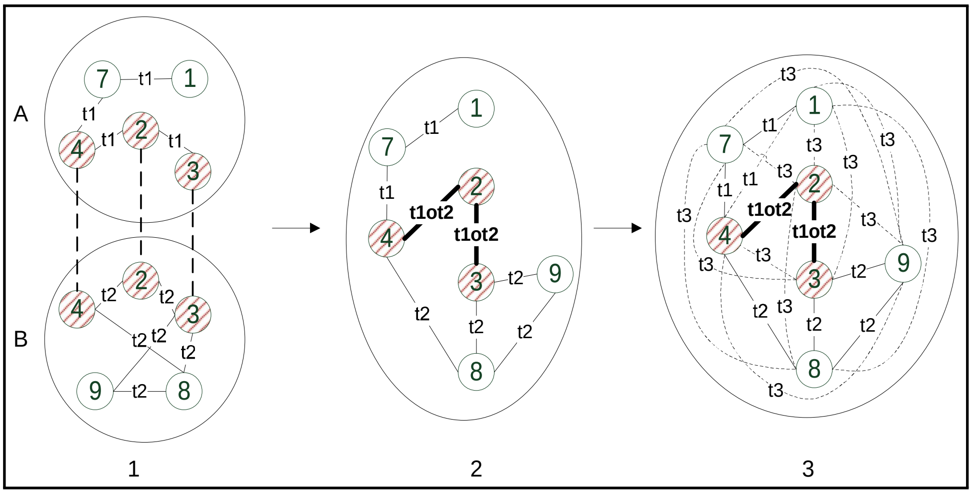


Figure 4.8: *CP* trust propagation strategy that explores true potential of consolidating MuDi social networks by evaluating indirect trust in consolidated graph. Pattern filled nodes represent users that are part of multiple social networks and bold ties represent aggregated information from both the networks consolidated.

4.7 Trust Propagation Algorithms

In the consolidated version of individual networks, the propagated trust between pairs of participants with a history of collaboration is available from the ties between them, but it needs to be calculated between cross-network participants or between same network users that we now believe to have been improved by the consolidation of information from multiple networks.

4.7.1 Strongest vs Shortest Path

There are different trust propagation algorithms discussed in Section 2.2.3, but here transitive decay-based trust from Walter et al. (2008) is evaluated for strongest and shortest paths of trust. There is transitivity of trust with a decay factor in social networks, and there are empirical studies which prove this point (Holland and Leinhardt, 1972; Leinhardt, 1972). People prefer to trust friends of their friend rather than strangers, but the strength of that trust decreases with an increase in the path length.

The decay of trust can be implemented in two ways. The first is the strongest path, in which trust decays yet is not controlled by the distance between individuals; rather, it takes place as a result of multiplication of trust values (in between [0,1]) between intermediate nodes in the path. The second uses the concept of the shortest path, which primarily sees trust decay with an increase in distance between individuals, and the multiplication of trust values over the shortest path generates the final trust value between them.

4.7.1.1 Strongest Path

This trust inference mechanism searches for the trust path with the highest value of trust, regardless of its length (the number of individuals involved in that path). Therefore it relies heavily on the transitivity of trust (see Section 2.2.3.3).

Pseudocode:

Pseudocode 10 Pseudocode that returns strongest trust value and length of that path for any two users s and t

```
procedure TRUSTCALCULATOR(s, source; t, target)
    trust ← { }
    pathLength ← { }
    visited ← {s:1}
    if t is adjacent to s then
        return s trust in t ← edgedata.get(weight)
    else:
        heapq.heappush(queue,(1, s))
        while queue do
            (trustValue,nodeURI) ← heapq.heappop(queue)
            trust[nodeURI] ← trustValue
            edata ← iter(g[nodeURI].items())
            for nghbrURI, edgedata in edata do
                nghbrnode_trust ← trust[nodeURI] * edgedata.get(weight)
                if nghbrURI not visited or nghbrnode_trust > visited[nghbrURI] then
                    visited[nghbrURI] ← nghbrnode_trust
                    if nodeURI in pathLength then
                        pathLength[nghbrURI] ← pathLength[nodeURI]+1
                    else:
                        pathLength[nghbrURI] ← 1
                    end if
                    heapq.heappush(queue, (nghbrnode_trust, nghbrURI))
                end if
            end for
        end while
    end if
    return trust[t], pathLength[t]
end procedure
```

Pseudocode 10 is an adapted version of the Dijkstra algorithm from the Python NetworkX API[1] which returns trust value and trust path length for the strongest path connection.

It takes source (s) and target (t) participants as its input parameter and returns trust value and path length that exists between them in the trust network. If s and t are adjacent nodes, it returns a trust rating on the link between them (*line 6*), otherwise it traverses the nodes in the network in a Dijkstra-like fashion, starting from the immediate neighbour of the s. If each of the neighbour nodes (*nghbrURI* at *line 13*) either has not been visited before or has trust greater than that obtained from any alternate path (*line 15*), then it is updated in the *visited dictionary* data structure (*line 16*). If the node is visited for the first time, its *pathLength* is set as 1 (*line 20*), otherwise the already recorded value is incremented by one (*line 18*). Then, it pushes the *nghbrURI* along with the calculated *nghbrnode_trust* in the heap to be traversed for its neighbours in the next iteration (*line 22*). This is repeated until it reaches the end of the network, visiting all the nodes. Finally, trust value and path length of the t node is returned to the calling function (*line 27*).

4.7.1.2 Shortest Path

The view that the shortest path trust path is the most accurate also holds that trust decays along paths between individuals in social networks, but that this decay depends primarily on the length of the trust path and not on the trust values between them. Hence, this mechanism first searches for the shortest of all paths, in terms of the number of ties between two nodes, then multiplies trust values associated with those edges to find the final trust value. If there are multiple paths of same shortest length, then the idea of the strongest path discussed in Section 4.7.1.1 is used to select one of the paths.

Pseudocode:

Pseudocode 11 implements this trust evalaution strategy and uses an extended version of the Breadth First Search algorithm from the Python NetworkX API[1] to find all the shortest paths that exist between users s and t.

First, analysing procedure *TrustCalculator* (in Pseudocode 11), and taking t as a reference, immediate neighbours to all the nodes in the network are extracted in a *dictionary* data structure (*pred*) with *keys* (in *pred*) representing nodes of the network and *values* (in *pred*) as the immediate neighbours connected to the *keys* (*line 7*). If it contains s as a *key*, then the algorithm is progressed (to *line 11*) due to existence of a path between s and t, otherwise it returns the message that t does not exists (or not reachable) in the network (*line 9*).

[1] https://networkx.github.io/

Pseudocode 11 Pseudocode that returns trust value for shortest trust path for any two users s and t

```
procedure ShortestPathTrustCalculator(s, source; t, target)
    strongestTrustValue ← [ ]
    trustValue ← 1
    valuelengthpair ←
    trustpaths = TrustCalculator(s, t)
    for path in trustpaths do
        plength ← 0
        for node in path do
            if node is not s then
                trustValue ← trustValue * edgedata.get(weight)
                plength ← plength + 1
            end if
        end for
        strongestTrustValue.append(trustValue)
        pathLengths[trustValue] ← plength
    end for
    return max(strongestTrustValue), pathLengths[max(strongestTrustValue)]
end procedure
procedure TrustCalculator(s, source; t, target)
    pred= { }
    trustPath ← [[s,0]]
    pthlength ← 1
    ind ← 0
    trustPaths ← [ ]
    pred ← determine neighbours of each node using BFS taking t as a reference
    if not pred.has_key(s) then
        return t is not reachable from s
    end if
    while ind >= 0 do
        nodeURI,i ← trustPath[ind]
        if t reached then
            trustPaths.append(trustPath)
        end if
        if len(pred[nodeURI]) > i then
            ind ← ind + 1
            if ind is equal to pthlength then
                trustPath.append([pred[nodeURI][i],0])
                pthlength ← pthlength + 1
            else
                trustPath[ind] ← [pred[nodeURI][i],0]
            end if
        else:
            ind ← ind - 1
            if ind >= 0 then
                trustPath[ind][1] ← trustPath[ind][1] + 1
            end if
        end if
    end while
    return trustPaths
end procedure
```

While loop (*line 11*) runs until there is any untraversed trust path left in the network. In this loop, if t is reached, list *trustpath* having a set of nodes in the path (between s and t) is added to the list of *trustpaths* (*line 13*), thus making a collection of all the shortest trust paths that exists between s and t.

If there are any neighbours of *nodeURI* (measured as *(len(pred[nodeURI])* at line 16), these are added to the *trustpath*, because they can possibly link to the t node (*line 19*). Otherwise, counter *ind* extracts already known neighbour of *nodeURI* (*line 22*) and moves to find another trust path.

If *nodeURI* has no neighbours, the algorithm is moved back towards the s node step by step decrementing the *ind* counter (*line 24*) to find alternate routes to the t node from any other predecessor node of the *nodeURI* that has neighbours to traverse.

At the end, this algorithm returns the list of all *trustPaths* (to procedure *Shortest-PathTrustCalculator*), that connects s and t through the shortest distances in terms of the number of the edges between them. Each of the shortest path is traversed one by one (*line 6* in procedure *ShortestPathTrustCalculator*), multiplying the trust values associated with the links in between (*line 10*), and an updated final trust value for each path is calculated. The length of the path having the highest final trust value is returned and the value is selected to be the trust value between users s and t (*line 17*).

4.7.2 Example

The two trust propagation algorithms presented are applied to the consolidated version of individual Networks A and B taken from Figure 4.8 for a pair of participants, *1* and *9*. Here, the ties between participants from these networks are also labelled with sample trust values, to better depict the reasoning behind selection of paths.

Analysing Figure 4.9 for the selection of the strongest trust path, there are three potential trust paths that are candidates for selection. Propagated trust over *Path 1* ($1 \rightarrow 7 \rightarrow 4 \rightarrow 8 \rightarrow 9$) is 0.02, over *Path 2* ($1 \rightarrow 7 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 9$) trust value results in 0.06 and 0.01 is the trust propagated between participants 1 and 9 over *Path 3* ($1 \rightarrow 7 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 8 \rightarrow 9$). Of these three paths, *Path 2* is the one that gives maximum trust between Participants 1 and 9, so according to the definition of the strongest trust path, 0.06 is the propagated trust value and *Path 2* is the resultant trust path with length of 5.

If Figure 4.9 is analysed to evaluate the trust value over shortest trust path, then again there are three potential trust paths. Path 1 ($1 \rightarrow 7 \rightarrow 4 \rightarrow 8 \rightarrow 9$) has the length of 4, *Path 2's* ($1 \rightarrow 7 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 9$) length is 5 and the length of *Path 3* ($1 \rightarrow 7 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 8 \rightarrow 9$) is 6. Using the definition of shortest path algorithm, *Path 1* is the selected trust path between Participants *1* and *9* and 0.02 is the trust value between them.

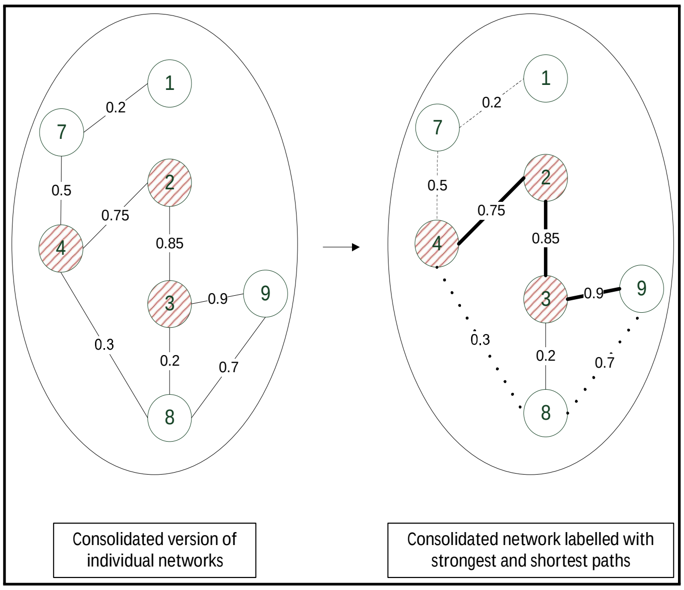


Figure 4.9: Sample consolidated version of individual networks *A* and *B* taken from Figure 4.8 analysed for strongest and shortest paths of trust. Dashed line path belongs to both strongest and shortest paths while thicked line and dotted line paths belong solely to strongest and shortest paths respectively.

Both these trust propagation algorithms are analysed using the simulation and real-world experiment to see which algorithm gives better values of trust.

4.8 Conclusions

This chapter discusses trust data fusion and inference techniques over multiple social networks and analyses the performance of the different data fusion techniques with three sample data records. The proposed trust aggregation function is defined as a bounded, idempotent and monotonic function that is able to distinguish *absence of trust* from the *distrust*. Furthermore, it should consider the importance of both the trust data and the source of that data. Naive methods of Summation (S), Weighted Average (WA) and Ordered Weighted Average (OWA) are unable to meet the aforementioned conditions, hence are apparently not appropriate choices for the overlapping trust aggregation.

Direct trust evaluations between users are preferred, while a decay-based approach is used for indirectly connected users. The two trust decay techniques discussed are the selection of the strongest and the shortest paths that consider the strength of the link and the length of the link as the measurement of trust, respectively. The *Consolidate Propagate (CP)* is the trust propagation strategy proposed for this study, which consolidates individual networks before evaluating trust for indirectly connected participants.

It is our hope that the proposed trust aggregation strategy will generate shorter high quality trust paths than those in individual networks.

In the next chapter, the discussed trust data fusion and propagation techniques will be used in a simulation to analyse the impact of consolidation of MuDi networks on a set of derived trust factors.

Chapter 5

Experiment I - Simulation Analysis of MuDi Trust Aggregation

5.1 Introduction

This chapter describes how the semantic infrastructure from Chapter 3 was used to run a simulation testing the proposed idea of MuDi trust consolidation on a sample set of trust networks. Pairs of networks were generated in series, using the proposed trust ontology for encoding trust and network information. In each network the generated random trust values were added to the links between sample participants, then the networks were merged and trust aggregation techniques used to consolidate multiple trust values between users. Results from different aggregation techniques were analysed in terms of the trust aggregation properties mentioned in Section 4.3, with the aim of identifying the optimum technique to use for real-world data.

5.2 Generating Sample MuDi Social Networks

In this experiment, pairs of social networks were generated with a varying percentage of overlapping nodes and overlapping ties in each pair of networks, so we could see the effect on trust values when networks with different percentages of overlap were consolidated.

We used the following vocabulary in the subsequent sections of this chapter: N1 and N2 represent the original networks being consolidated, MuDi is the final consolidated network; and CN1 and CN2 are sub-networks in the MuDi that represent the original networks. *PO* stands for percentage of Participant Overlap and *TO* means percentage of Tie Overlap.

5.2.1 Wiring Links and Adding Trust Values on the Links

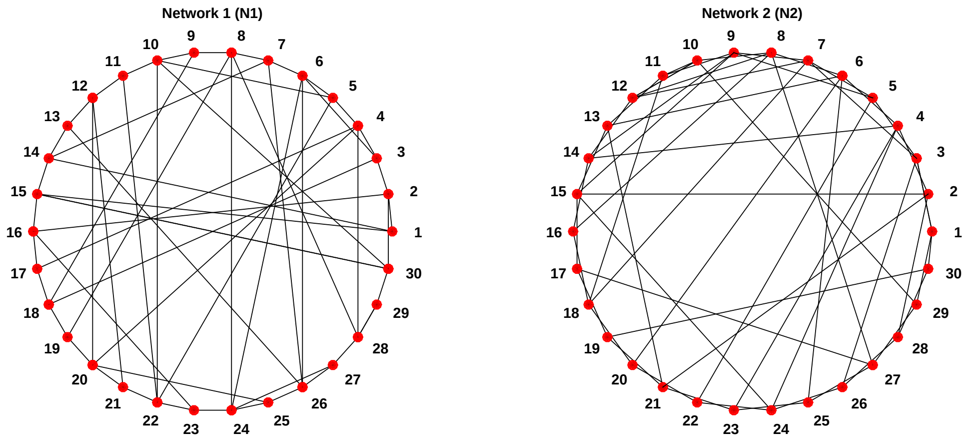
This section extends the network generation model described in Section 2.3.2 for generating a pair of networks with overlapping participants and ties. We used a three-step mechanism for wiring connections between participants to achieve a given percentage of PO and TO.

1. First, users in each of the constituent networks were connected in a ring lattice to ensure the connectedness of the networks. If there was no TO between the networks, then networks were still connected but overlapping connections were replaced with non-overlapping connections.

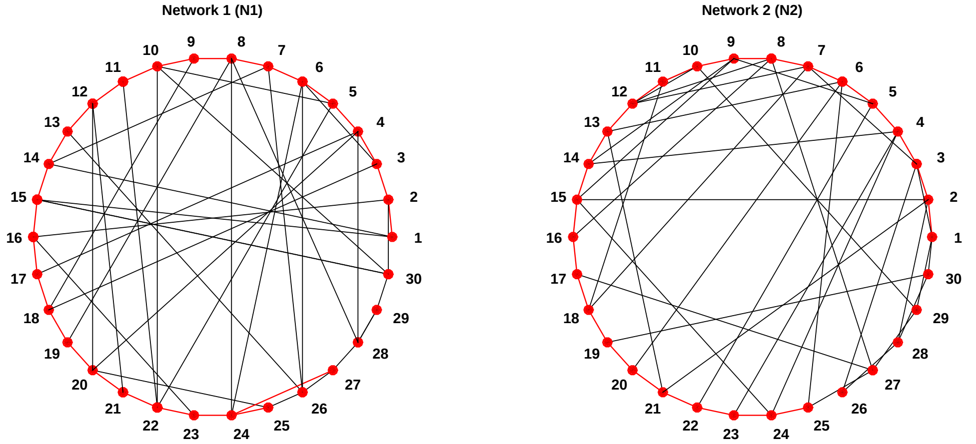
2. Then, the remaining percentage of overlapping ties was created between randomly selected pair of participants with probability $1/N_p$ among all present in the network, where N_p represents a pair of participants.

3. In the final step, non-overlapping ties between the random pair of participants were created with the same probability as that in Step 2. Duplicate ties were forbidden in Steps 2 and 3 to ensure the target density of the networks.

We assumed that the networks contained trust information on their ties, represented as a continuous value in the range $(0, 1)$. The ties of the generated sample networks were randomly labelled with these values and represented subjective asymmetric trust between users, with values near to 0 representing low trust and those close to 1 high trust between connected individuals.



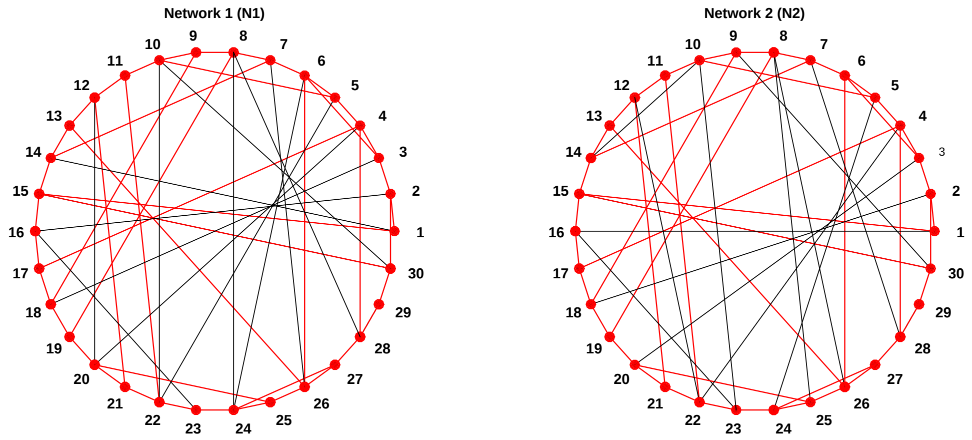
(a) PO=100%, TO=0%



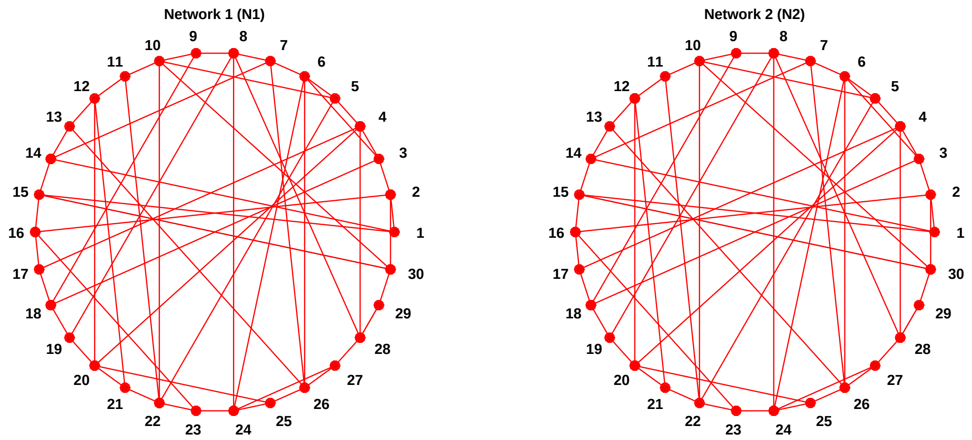
(b) PO=100%, TO=40%

Figure 5.1: Two randomly generated connected sample social networks N1 and N2 with PO = 100% and TO = [0%,40%]. Trust values on the links are intentionally missed out just to keep the diagram clean and simple. Red represents overlapping portions between networks.

Figures 5.2 and 5.3 demonstrate the three step procedure for generating a sample pair of networks for a fixed value of PO (i.e. $PO = 100\%$) and varying value of TO (i.e. for $TO = 0\%$, 20%, 80%, 100%). All the users in both the networks were overlapping, represented as nodeID in the range [1, 30]. At $TO = 0\%$, ties between user pairs in both the networks are not alike at all, meaning that none of the user pairs were connected in both the networks. As a result, users in sample Network 1 (N1) were connected as a connected ring with immediate neighbours, while those connections in sample Network 2 (N2) were replaced with alternate neighbours. However, with an increase in TO, ties started to overlap (shown in *red*) and, at TO = 40%, both N1 and N2 were connected in a ring with 40% of the ties overlapping. At TO = 80%, pattern of ties between user pairs in both the networks started to appear similar. At PO = 100%, TO = 100%, all the users and the ties between them were overlapping and hence both Networks N1 and N2 were identical.



(a) PO=100%, TO=80%



(b) PO=100%, TO=100%

Figure 5.2: Two randomly generated connected sample social networks N1 and N2 with PO = 100% and TO = [80%,100%]. Trust values on the links are intentionally missed out just to keep diagram clean and simple. Red represents overlapping portions between networks.

5.2.2 Ensuring Small World Properties

In this simulation, we focused on the concept of consolidating professional social networks and it is important that the networks we generated have the characteristics of real-world professional networks (as described in Table 2.1 - C needed to be in the range [0.34, 0.6] and L in [4.92, 7.57]). Using this stipulation as a guide we conducted two pre-experiments to find the value of Density (D) and Number of nodes (N) to ensure that values of these properties lay in a comparable range.

First experiment aimed to find the value of D that generated a clustering coefficient (C) and average length of shortest paths (L) within the range specified in Table 2.1. Table 5.1 shows the parameters used for this experiment, along with a description. Here, N, *PO* and *TO* are purposely fixed at 60, 60% and 40% respectively, with D varying in the range [0.05, 0.85] to measure the behaviour of C and L in the generated networks N1 and N2.

Table 5.1: Experiment parameters along with their description to select the value of Density *D*.

Parameters	Description
No of Nodes (N)	60
Density (D)	$[0.05, 0.85]$
PO	60%
TO	40%

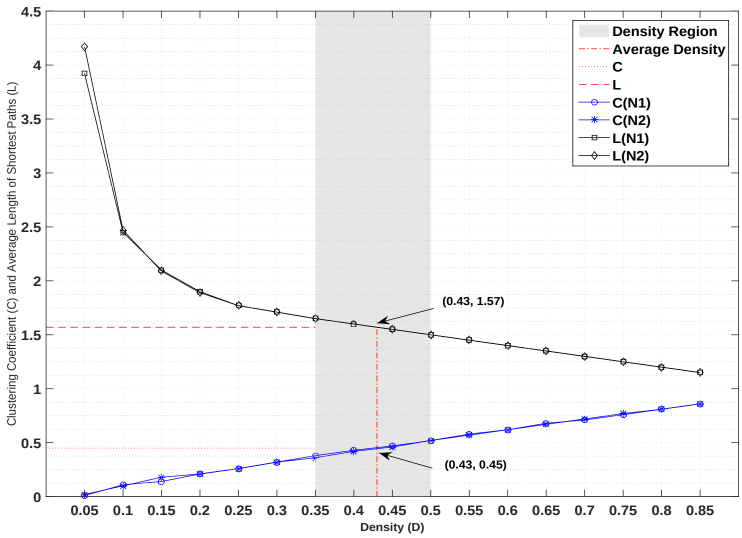


Figure 5.3: Values of C and L from example networks N1 and N2 are shown with the varying value of D. The two pairs of coordinates show the selected value of D and the corresponding value of C and L.

Figure 5.3 shows the results of this experiment. For D in the range of [0.35,0.5] (depicted as a shaded region), C resided within the target range, that is, [0.35,0.56], however L appeared to be smaller (i.e. 1.57) than the value given in Table 2.1 and its maximum value was near to that in the literature was 4.17, but it had a very low C (i.e. 0.01). The reason for the low value of L is the presence of a greater number of short connections, inevitable with the high value of D. So with the designed settings of this simulation, there was a trade-off between C and L, bearing in mind the value of D. To ensure the target value of C, we opted for C lying exactly in the range mentioned in Table 2.1 with a comparatively low value of L.

Taking the average of density values on the edges of the shaded area (i.e. $(0.35+0.5)/2$) it generated a D of 0.43, and this was used in the next pre-experiment and later in experiments relating to trust measurement.

To select the value of N for the subsequent simulation experiment, the second pre-experiment aimed to find a value of N that gave the same values of C and L for both the sample generated networks N1 and N2 to be consolidated, to make sure that both the generated networks had the same small-world characteristics. Similar to the last experiment, three of the parameters D, *PO* and *TO* were intentionally fixed at 0.43, 60% and 40% respectively, while N varied over the range [5, 85] to find the value of N that ensured ratio of 1 for C and L between N1 and N2.

Table 5.2: Experiment parameters along with their description to select the value of number of Nodes *N*.

Parameters	Description
No of Nodes (N)	$[5, 85]$
Density (D)	0.43
PO	60%
TO	40%

The results of the experiment in Figure 5.4 show that initially, for $N < 30$, C and L were not alike from Networks N1 and N2, due to the lesser number of connections due to the fixed D. But they started to converge and became similar at $N = 30$ (depicted as a *red* vertical line), which held true for the rest of the data. Hence, $N = 30$ is the minimum value that could be selected as number of nodes for the simulation. The reason for selecting $N = 30$ and not 40 or 60 (as C and L are same for these values) was to keep the network generation and trust computation costs low.

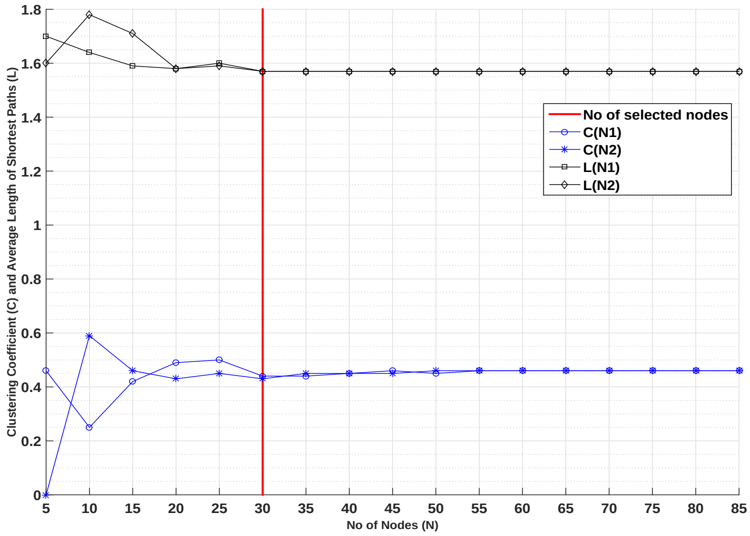


Figure 5.4: Values of C and L for example networks N1 and N2 are shown with the varying value of N. Vertical red line represents the selected value of N = 30

5.3 Experiment Design

To analyse the impact of consolidating networks on trust properties, a simulation experiment was designed that varied several parameters to establish a consolidation technique that satisfied the trust aggregation properties (described in Section 4.3) and later could be used with real-world data for making trust computations.

The design of the experiment included two portions; the first described different network and consolidation parameters input into the simulation, for example, *PO*, *TO*, N and so on. The second set was those used for evaluating the performance of consolidating MuDi social networks, such as average strength of trust ties and average length of trust paths (further described in Section 5.3.2).

This experiment was based on a numerical simulation of the consolidation of pairs of networks, using Summation (S), Weighted Average (WA), Weighted Ordered Weighted Averaging (WOWA) and Induced Ordered Weighted Averaging (IOWA) techniques. It was implemented in Python (including the code for network generation, consolidation and trust inference algorithms), and the NetworkX[1] API was used for measuring network properties.

[1] http://networkx.github.io/documentation/latest/reference/introduction.html

5.3.1 Simulation Parameters

When conducted using real-world data, consolidation of networks would generally uses a heuristic approach of meta-data comparison to identify participants that appear in both networks (for example, by comparing *familyName* and *givenName* properties). The result will be a certain percentage of overlapping participants (represented as PO) with potentially overlapping ties (referred as TO) between them.

The simulation parameters in this experiment were of two types. The first type of parameters was network parameters, for example, N, D, C and L, already discussed in Section 5.2.2. The second type was consolidation parameters such as PO, TO and so on. The network parameters for this experiment were pre-determined by running two pre-experiments, but values of consolidation parameters needed to be found. Based on the advantages and disadvantages of different aggregation techniques discussed in Section 4.5, it was clear that the WA technique considered the importance of trust sources but did not help in differentiating the importance of data from different trust data sources. Accordingly, values of p_{WA} were set as [0.5, 0.5] to emphasise that both trust sources were of equal importance. IOWA permitted us to do this using the p vector; for example, if trust values from network N1 were considered to be more trustworthy than N2, it could be achieved by setting the parameter p_{IOWA}= [N1, N2]. The reliability vector for trust data could also be altered accordingly, i.e. w_{IOWA} = [0.8, 0.2]. Unlike either of the above techniques, WOWA permuted the trust values in $a_\sigma(1) \leq a_\sigma(2) \leq \ldots \leq a_\sigma(n)$ order and allowed us to set the importance of high trust values (i.e. w_{WOWA} = [1, 0.5]) and their sources, (i.e. p_{WOWA} = [0.8, 0.2]) as high, compared to low values and their sources.

In our simulation we wanted to show the effect on the trust properties (discussed in Section 5.3.2), since PO and TO vary. However, TO is constrained by PO as, to achieve a certain percentage of TO, at least $PO \geq TO$[2] should be in place, otherwise the number of overlapping connections would exceed the maximum possible number of ties between the subsets of overlapping participants. For example, a 100% TO was only possible if there was 100% PO as well.

We ran a number of simulations, setting PO at 40%, 60%, 80%, and 100%, and for each setting of PO (except 40%PO[2]) allowed TO to vary from 0 to PO in increments of 20%. Then in each simulation we consolidated the trust information on the links using S, WA, WOWA and IOWA. Table 5.3 shows the values and ranges for all the parameters of the simulation.

[2]For 40%PO, $PO > TO$ should be true, because 40%PO = 40% ∗ 30 = 12, the maximum possible number of undirected ties between overlapping participants can be $(12 * 11)/2 = 66$ and 40%TO = $(40\% * (0.43 * (30 * 29))/2) = 74$. So the required number of overlapping ties 74 exceeds maximum possbile number of ties ties 66.

Table 5.3: Network and consolidation parameters used for this study

Network Parameters	Description
N	30
D	0.43
C	0.45 ± 0.02
L	1.57
Ratio of C, D, L between $N1$ and $N2$	1
Consolidation Parameters	**Description**
PO	$[40\%, 100\%]$
TO	$[0, PO]$[2]
p_{WA}	$[0.5, 0.5]$
p_{IOWA}	$[N1, N2]$
w_{IOWA}	$[0.8, 0.2]$
p_{WOWA}	$[0.8, 0.2]$
w_{WOWA}	$[1, 0.5]$

5.3.2 Selecting Trust Properties for Evaluation

To quantify the performance of this trust inference mechanism on consolidated MuDi networks and CN1, CN2 (sub-networks that represent N1 and N2 in consolidated MuDi networks), there are two variables identified from the literature; namely *strength of trust ties* and *length of trust paths*. $T_{a_i,\ldots,a_j}$ is the *strength of the trust tie* between participants a_i and a_j, and it shows the amount of trust a_i holds in a_j, and the *length of the trust path* is the number of ties involved in the $path(a_i, a_j)$ (Equation 2.10 describes their relation). As direct interaction history is prioritised over recommendations, $path(a_i, a_j)$ = 1 for directly connected participants, while it is evaluated for indirectly connected participants using algorithms, as mentioned in Section 4.7. Approximate estimations of these two trust properties were made for each of the networks N1, N2, CN1, CN2 and MuDi by taking the average of the trust estimations for each pair of participants in the network and examining them for the change as a result of consolidation. For example, if TS, TL represents such values for *strength of ties* and *length of trust paths* respectively, then it can be calculated as described in Equations 5.1 and 5.2:

$$\text{TS} = \frac{1}{n(n{-}1)} \sum_{a_i, a_j \epsilon N_p} T_{a_i,\ldots,a_j} \tag{5.1}$$

$$\text{TL} = \frac{1}{n(n{-}1)} \sum_{a_i, a_j \epsilon N_p} path(a_i, a_j) \tag{5.2}$$

We were looking for a consolidation that did not damage existing trust properties, but used the additional information to enhance them. In terms of TS and TL this meant that we would like TS for CN1, CN2 and MuDi to remain close to that of N1, N2, even if there were no significant *PO* and *TO*. If TS was maintained in this way it would show that damage was minimised during consolidation. Furthermore, we would expect TL to decrease significantly overall due to the emergence of additional trust paths stronger than those in N1 and N2. If TL decreased it would show that the consolidation has successfully enhanced trust calculations by opening up new trust paths.

5.4 Results and Analysis

There were two types of trust measurements in the networks. The first, on direct ties, was aggregated as a result of merging the individual networks. The second type of measurements was evaluated for each pair of indirectly connected participants N_P in (N1, N2), (CN1, CN2) and consolidated MuDi networks separately, and these values depended on the trust between intermediate nodes in the trust path in that specific network. As the Density, D, of network in each of the constituent networks is 0.43, this meant each original network N1 and N2 had 43% direct connections and that 57% of the trust evaluations are based on finding trust paths.

Hypothesis: This experiment examines the second claim of the hypothesis (Section 1.3) which states that the data fusion techniques allow us to aggregate trust metrics from MuDi social networks and respect the integrity of trust from individual networks, while opening up many additional trust paths.

5.4.1 Trust data Description

There were two types of data available for two different trust propagation algorithms (already discussed in Section 4.7) in this simulation. The first was the average strength of trust ties, represented using the TS metric for varying values of *PO* and *TO*, and the second was the average length of the trust paths, shown as the TL metric for varying value of *PO* and *TO*.

5.4.1.1 Impact of varying Participant Overlap (PO) and Tie Overlap (TO) on average strength of trust ties (TS)

Tables 5.4 and 5.5 present the values of TS for varying consolidation parameters *PO* and *TO* for the strongest and shortest path recommendation algorithms, and two sample results (for $PO = [60\%, 100\%]$) for each of the technique are depicted in Figure 5.5. For different percentages of *PO* and *TO*, TS was evaluated for N1, N2 and then evaluated for CN1, CN2 and the MuDi network for each aggregation strategy (S, WA, WOWA, IOWA).

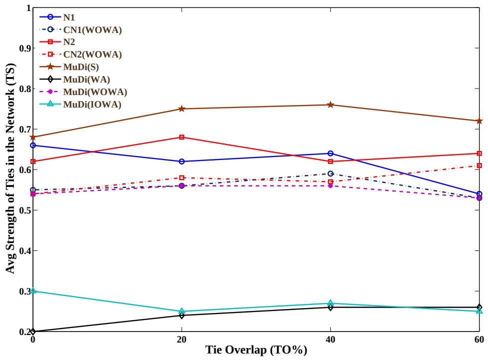
First, if we look at TS for (CN1, CN2) in Table 5.4 and Figures 5.5(a), 5.5(b) (for strongest path recommendation), it can be seen that the value of this metric for WOWA resides in between two extreme approaches S and WA. S amplifies the trust by summing the values available on the ties, hence inflating the trust up to (0.94, 0.94) at $[100\%PO, 100\%TO]$, while WA dampens trust down to (0.21, 0.21) at $[40\%PO, 0\%TO]$, which were (0.63, 0.60) and (0.64, 0.65) in (N1, N2) respectively. WOWA, in both of the above mentioned cases calculates more stable metric with values of (0.69, 0.69) and (0.53, 0.54) respectively. However, the fourth technique, IOWA, shows somewhat different to the other approaches due to prioritising trust data source N1 over N2. For N1, although it shows comparatively high trust values near to WOWA (i.e. 0.50 at $[40\%PO, 0\%TO]$ in CN1), for N2 it deteriorates the trust metrics and makes them even worse than those given by WA (i.e. 0.18 at $[40\%PO, 0\%TO]$ in CN2).

Table 5.4: Average strength of ties (TS) for four different trust aggregtion mechanisms using strongest path recommendation algorithm i.e. Summation (S), Weighted Average (WA), Weighted Ordered Weighted Averaging (WOWA) and Induced Ordered Weighted Averaging (IOWA) in the networks with varying percentage of *PO* and *TO*. CN1 and CN2 represent sub-networks in the consolidated MuDi networks. Data points highlighted in **bold** are depicted in Figures 5.5(a) and 5.5(b).

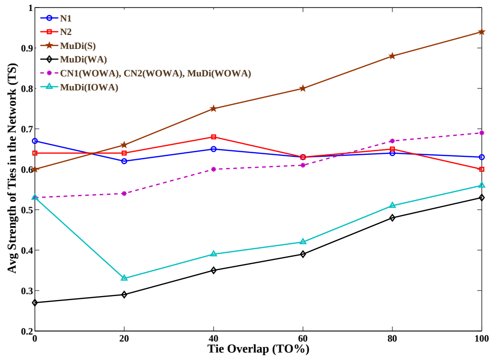
PO	*TO*	Avg. Strength of Ties in the Network (TS)													
		N1	N2	CN1				CN2				MuDi			
				S	WA	WOWA	IOWA	S	WA	WOWA	IOWA	S	WA	WOWA	IOWA
40	0	0.64	0.65	0.66	0.21	0.53	0.50	0.67	0.21	0.54	0.18	0.68	0.18	0.53	0.23
	20	0.59	0.62	0.69	0.25	0.52	0.41	0.73	0.27	0.56	0.15	0.71	0.20	0.52	0.21
	30	0.63	0.60	0.73	0.27	0.56	0.45	0.72	0.26	0.54	0.15	0.71	0.21	0.52	0.22
60	0	**0.66**	**0.62**	0.66	0.23	**0.55**	0.54	0.66	0.23	**0.54**	0.29	**0.68**	**0.20**	**0.54**	**0.30**
	20	**0.62**	**0.68**	0.73	0.28	**0.56**	0.41	0.76	0.28	**0.58**	0.20	**0.75**	**0.24**	**0.56**	**0.25**
	40	**0.64**	**0.62**	0.78	0.32	**0.59**	0.46	0.75	0.31	**0.57**	0.22	**0.76**	**0.26**	**0.56**	**0.27**
	60	**0.54**	**0.64**	0.71	0.30	**0.53**	0.41	0.81	0.35	**0.61**	0.24	**0.72**	**0.26**	**0.53**	**0.25**
80	0	0.64	0.62	0.63	0.24	0.53	0.54	0.63	0.25	0.54	0.39	0.65	0.23	0.54	0.37
	20	0.69	0.64	0.74	0.30	0.59	0.41	0.72	0.29	0.57	0.28	0.74	0.28	0.58	0.32
	40	0.65	0.68	0.80	0.35	0.61	0.44	0.80	0.36	0.61	0.32	0.80	0.32	0.60	0.34
	60	0.65	0.62	0.81	0.41	0.62	0.49	0.82	0.42	0.62	0.37	0.79	0.36	0.60	0.37
	80	0.59	0.60	0.78	0.40	0.58	0.47	0.85	0.42	0.63	0.38	0.77	0.34	0.56	0.35
100	0	**0.67**	**0.64**	0.60	0.27	**0.53**	0.53	0.60	0.27	**0.53**	0.53	**0.60**	**0.27**	**0.53**	**0.53**
	20	**0.62**	**0.64**	0.66	0.29	**0.54**	0.33	0.66	0.29	**0.54**	0.33	**0.66**	**0.29**	**0.54**	**0.33**
	40	**0.65**	**0.68**	0.75	0.35	**0.60**	0.39	0.75	0.35	**0.60**	0.39	**0.75**	**0.35**	**0.60**	**0.39**
	60	**0.63**	**0.63**	0.80	0.39	**0.61**	0.42	0.80	0.39	**0.61**	0.42	**0.80**	**0.39**	**0.61**	**0.42**
	80	**0.64**	**0.65**	0.88	0.48	**0.67**	0.51	0.88	0.48	**0.67**	0.41	**0.88**	**0.48**	**0.67**	**0.51**
	100	**0.63**	**0.60**	0.94	0.53	**0.69**	0.56	0.94	0.53	**0.69**	0.56	**0.94**	**0.53**	**0.69**	**0.56**

Table 5.5: Average strength of ties (TS) for four different trust aggregtion mechanisms using shortest path recommendation algorithm i.e. Summation (S), Weighted Average (WA), Weighted Ordered Weighted Averaging (WOWA) and Induced Ordered Weighted Averaging (IOWA) in the networks with varying percentage of *PO* and *TO*. CN1 and CN2 represent sub-networks in the consolidated MuDi networks. Data points highlighted in **bold** are depicted in Figures 5.5(c) and 5.5(c).

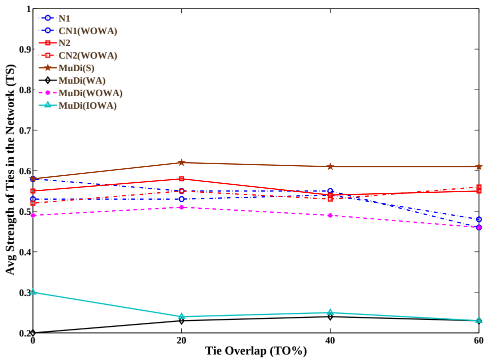
PO	*TO*	Avg. Strength of Ties in the Network (TS)													
		N1	N2	CN1				CN2				MuDi			
				S	WA	WOWA	IOWA	S	WA	WOWA	IOWA	S	WA	WOWA	IOWA
40	0	0.53	0.55	0.56	0.21	0.48	0.44	0.58	0.21	0.50	0.17	0.52	0.17	0.44	0.22
	20	0.52	0.55	0.59	0.24	0.47	0.39	0.64	0.26	0.52	0.15	0.54	0.19	0.44	0.19
	30	0.55	0.52	0.66	0.26	0.52	0.42	0.61	0.25	0.49	0.14	0.55	0.19	0.44	0.20
60	0	**0.58**	**0.55**	0.61	0.23	**0.53**	0.51	0.60	0.23	**0.52**	0.28	**0.58**	**0.20**	**0.49**	**0.30**
	20	**0.55**	**0.58**	0.65	0.27	**0.53**	0.39	0.67	0.27	**0.55**	0.19	**0.62**	**0.23**	**0.51**	**0.24**
	40	**0.55**	**0.54**	0.68	0.30	**0.54**	0.42	0.66	0.29	**0.53**	0.21	**0.61**	**0.24**	**0.49**	**0.25**
	60	**0.46**	**0.55**	0.63	0.29	**0.48**	0.37	0.73	0.32	**0.56**	0.23	**0.61**	**0.23**	**0.46**	**0.23**
80	0	0.56	0.54	0.59	0.24	0.51	0.50	0.59	0.25	0.52	0.38	0.59	0.23	0.51	0.38
	20	0.61	0.56	0.69	0.29	0.57	0.39	0.67	0.29	0.55	0.27	0.67	0.27	0.55	0.30
	40	0.57	0.60	0.74	0.34	0.59	0.42	0.75	0.34	0.59	0.30	0.71	0.31	0.57	0.32
	60	0.55	0.54	0.76	0.38	0.57	0.45	0.76	0.38	0.57	0.34	0.70	0.32	0.53	0.33
	80	0.50	0.53	0.74	0.38	0.55	0.44	0.78	0.39	0.58	0.37	0.69	0.32	0.51	0.32
100	0	**0.58**	**0.56**	0.59	0.27	**0.53**	0.53	0.59	0.27	**0.53**	0.53	**0.59**	**0.27**	**0.53**	**0.53**
	20	**0.54**	**0.53**	0.64	0.29	**0.54**	0.32	0.64	0.29	**0.54**	0.32	**0.64**	**0.29**	**0.54**	**0.32**
	40	**0.55**	**0.59**	0.72	0.34	**0.59**	0.37	0.72	0.34	**0.59**	0.37	**0.72**	**0.34**	**0.59**	**0.37**
	60	**0.52**	**0.55**	0.77	0.38	**0.59**	0.40	0.77	0.38	**0.59**	0.40	**0.77**	**0.38**	**0.59**	**0.40**
	80	**0.57**	**0.58**	0.87	0.46	**0.65**	0.48	0.87	0.46	**0.65**	0.48	**0.87**	**0.46**	**0.65**	**0.48**
	100	**0.58**	**0.52**	0.93	0.50	**0.66**	0.53	0.93	0.50	**0.66**	0.53	**0.93**	**0.50**	**0.66**	**0.53**



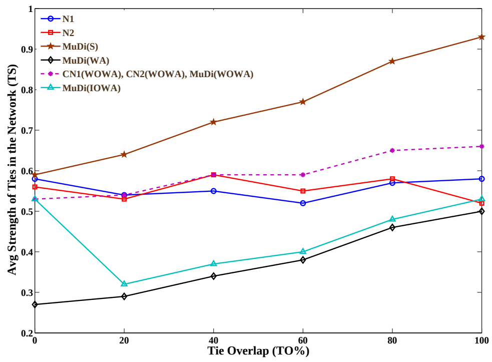
(a) Participant Overlap(PO) = 60%



(b) Participant Overlap(PO) = 100%



(c) Participant Overlap(PO) = 60%



(d) Participant Overlap(PO) = 100%

Figure 5.5: Sample values of TS selected from Tables 5.4 and 5.5 for depiction. N1 and N2 represent original networks, MuDi(S), MuDi(WA), MuDi(WOWA) and MuDi(WOWA) represent TS in MuDi for S , WA, WOWA and IOWA trust aggregation techniques. CN1 (WOWA) and CN2 (WOWA) show TS metric from sub-networks CN1 and CN2 in MuDi for WOWA aggregation technique. Figures 5.5(a) and 5.5(b) are from strongest path algorithm while Figures 5.5(c) and 5.5(d) are from shortest path algorithm.

Similarly, the TS metric for (CN1, CN2) in Table 5.5 and Figures 5.5(c), 5.5(d) (for shortest path recommendation) shows the same trend of results as of the strongest path recommendation and WOWA again resides between two extreme values, S and WA. Unlike the previous case, here all the trust metrics are somewhat dampened due to taking shortest path (with respect to the number of ties), rather than the strongest path (with respect to the strength of ties). For S it results in (0.93, 0.93) at $[100\%PO, 100\%TO]$ which were (0.94, 0.94) in the above scenario, and for WA it results in (0.21, 0.21) at $[40\%PO, 0\%TO]$ which were (0.21, 0.21) in the above case. Due to this dampening, however, TS for WOWA draws closer to N1 and N2 than those from the strongest path and results in (0.66, 0.66) (at $[100\%PO, 100\%TO]$) and (0.48, 0.50) (at $[40\%PO, 0\%TO]$) with respect to (0.58, 0.52) and (0.53, 0.55) from (N1, N2) respectively. As IOWA only respects trust metrics from one of the networks, N1, as in the earlier case its evaluated TS value is comparatively close to N1 for all values of PO and TO (i.e. 0.53 in comparison with 0.58 at $[100\%PO, 100\%TO]$ and 0.44 in comparison with 0.53 at $[40\%PO, 0\%TO]$), but is badly distorted for low PO and TO in N2 (i.e. 0.17 in comparison with 0.55 at $[40\%PO, 0\%TO]$).

Behaviour of the TS metric in MuDi (for strongest path recommendation) is also in accordance with those of CN1 and CN2, and the results of WOWA again appear to be more stable than those of the other techniques. S escalates the trust up to 0.94 at $[100\%PO, 100\%TO]$, while WA reduces it to 0.18 at $[40\%PO, 0\%TO]$, but WOWA maintains it at 0.69 and 0.53 respectively. IOWA at $[100\%PO, 100\%TO]$, appears to be a reasonable approach with TS of 0.56, but at $[40\%PO, 0\%TO]$, its distortion of trust values is similar to WA (i.e. TS = 0.23), so becomes an inappropriate aggregation choice when compared to WOWA.

The TS metric using WOWA approach (for shortest path recommendation) in MuDi also ends up having more stable measurements than the other approaches - S, WA and IOWA - without distorting trust metrics at any stage of the PO and TO. S and WA just distorts TS metric drastically at two different extremes (i.e. 0.93 at $[100\%PO, 100\%TO]$ and 0.17 at $[40\%PO, 0\%TO]$ respectively), while IOWA also performs poorly for low PO and TO. Only WOWA generates more stable metrics (0.66 at $[100\%PO, 100\%TO]$ and 0.44 at $[40\%PO, 0\%TO]$), and thus appears to be the best approach for trust aggregation (also seen in Figures 5.5(c) and 5.5(d)).

5.4.1.2 Impact of varying Participant Overlap (PO) and Tie Overlap (TO) on average length of trust paths (TL)

Tables 5.6 and 5.7 show TL metric for strongest and shortest path algorithms (respectively) for varying PO and TO. Two samples (for $PO = [60\%, 100\%]$) for each of the technique are depicted in Figure 5.6. Tabular data show values of TL for both N1 and N2, sub-networks CN1 and CN2 and consolidated MuDi version of original networks, while Figure 5.6 shows N1 and N2 along with CN1 and CN2 from WOWA while MuDi values from all the aggregation techniques (S, WA, WOWA and IOWA).

If the sub-networks CN1 and CN2 are considered for average length of trust paths TL (for strongest path recommendation), from Table 5.6 and Figures 5.6(a), 5.6(b), it is observed that the WOWA decreases the TL metric for all values of PO and TO in CN1 and CN2. For S, TL only decreases for $TO \geq 40\%$, whereas WA reduces it significantly for $TO \leq 40\%$ as well. IOWA also decreases TL for all values of PO and TO and for low PO it even gives better results then WOWA. Although reduction in TL using WOWA is less than either WA and IOWA, if coupled with TS metric, then the marginally smaller decrease in TL for WOWA is acceptable with the significantly better preservance of trust values in TS compared to WA and IOWA.

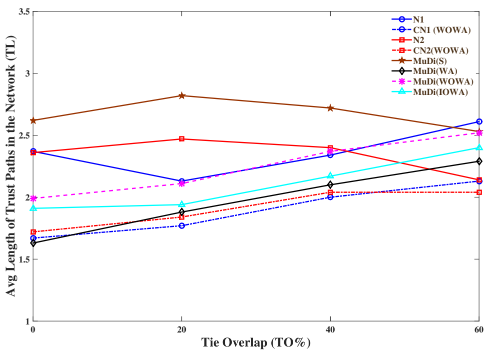
Table 5.6: Average length of trust paths (TL) for four different trust aggregtion mechanisms using strongest path recommendation algorithm i.e. Summation (S), Weighted Average (WA), Weighted Ordered Weighted Averaging (WOWA) and Induced Ordered Weighted Averaging (IOWA) in the networks with varying percentage of *PO* and *TO*. CN1 and CN2 represent sub-networks in the consolidated MuDi networks. Data points highlighted in bold are plotted in Figures 5.6(a) and 5.6(b)

PO	*TO*	Avg. Length of Trust Paths (TL)													
		N1	N2	CN1				CN2				MuDi			
				S	WA	WOWA	IOWA	S	WA	WOWA	IOWA	S	WA	WOWA	IOWA
40	0	2.68	2.41	2.41	1.56	2.01	2.20	2.29	1.55	1.89	2.03	3.04	1.82	2.40	2.26
	20	2.26	2.43	2.34	1.85	2.09	1.90	2.32	1.80	1.95	1.89	2.95	2.16	2.45	2.26
	30	2.31	2.39	2.33	1.80	1.94	1.93	2.61	1.83	2.03	1.89	3.17	2.16	2.45	2.29
60	0	**2.37**	**2.36**	2.06	1.45	**1.67**	1.82	2.18	1.45	**1.72**	1.99	**2.62**	**1.63**	**1.99**	**1.91**
	20	**2.13**	**2.47**	2.22	1.69	**1.77**	1.68	2.35	1.68	**1.84**	1.77	**2.82**	**1.88**	**2.11**	**1.94**
	40	**2.34**	**2.40**	2.13	1.87	**2.0**	1.86	2.22	1.86	**2.04**	1.97	**2.72**	**2.10**	**2.37**	**2.17**
	60	**2.61**	**2.14**	2.10	1.91	**2.13**	2.0	2.0	1.97	**2.04**	2.01	**2.53**	**2.29**	**2.52**	**2.40**
80	0	2.35	2.49	1.86	1.35	1.55	1.75	1.72	1.32	1.48	1.67	2.04	1.43	1.68	1.67
	20	2.37	2.29	1.98	1.48	1.59	1.60	2.03	1.48	1.64	1.59	2.29	1.59	1.77	1.72
	40	2.24	2.33	1.89	1.65	1.74	1.70	1.83	1.70	1.71	1.73	2.13	1.79	1.89	1.85
	60	2.60	2.29	1.96	2.07	2.01	1.98	1.87	2.06	1.92	1.95	2.20	2.22	2.17	2.16
	80	2.53	2.19	1.86	1.94	2.04	2.04	1.93	1.91	2.07	2.02	2.27	2.21	2.43	2.37
100	0	**2.39**	**2.52**	1.37	1.14	**1.20**	1.20	1.37	1.14	**1.20**	1.20	**1.37**	**1.14**	**1.20**	**1.20**
	20	**2.23**	**2.80**	1.70	1.27	**1.32**	1.33	1.70	1.27	**1.32**	1.33	**1.70**	**1.27**	**1.32**	**1.33**
	40	**2.58**	**2.37**	1.66	1.48	**1.49**	1.52	1.66	1.48	**1.49**	1.52	**1.66**	**1.48**	**1.49**	**1.52**
	60	**2.62**	**2.31**	1.62	1.60	**1.62**	1.64	1.62	1.60	**1.62**	1.64	**1.62**	**1.60**	**1.62**	**1.64**
	80	**2.20**	**2.24**	1.58	1.75	**1.74**	1.83	1.58	1.75	**1.74**	1.83	**1.58**	**1.75**	**1.74**	**1.83**
	100	**2.15**	**2.33**	1.64	1.88	**1.98**	1.89	1.64	1.88	**1.98**	1.89	**1.64**	**1.88**	**1.98**	**1.89**

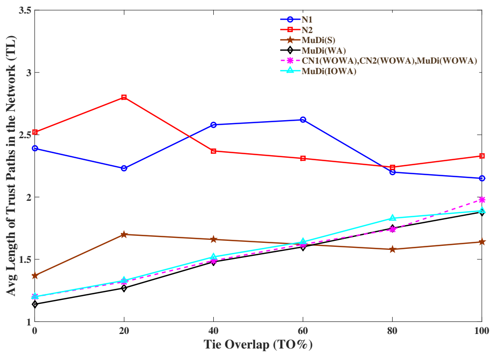
Table 5.7: Average length of trust paths (TL) for four different trust aggregtion mechanisms using shortest path recommendation algorithm i.e. Summation (S), Weighted Average (WA), Weighted Ordered Weighted Averaging (WOWA) and Induced Ordered Weighted Averaging (IOWA) in the networks with varying percentage of *PO* and *TO*. CN1 and CN2 represent sub-networks in the consolidated MuDi networks. Data points highlighted in bold are plotted in Figures 5.6(c) and 5.6(d).

PO	*TO*	Avg. Length of Trust Paths (TL)													
		N1	N2	CN1				CN2				MuDi			
				S	WA	WOWA	IOWA	S	WA	WOWA	IOWA	S	WA	WOWA	IOWA
40	0	1.57	1.57	1.52	1.52	1.52	1.52	1.50	1.50	1.50	1.50	1.69	1.69	1.69	1.69
	20	1.57	1.57	1.56	1.56	1.56	1.56	1.54	1.54	1.54	1.54	1.76	1.76	1.76	1.76
	30	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.57	1.79	1.79	1.79	1.79
60	0	**1.57**	**1.57**	1.44	1.44	**1.44**	1.44	1.43	1.43	**1.43**	1.43	**1.57**	**1.57**	**1.57**	**1.57**
	20	**1.57**	**1.57**	1.48	1.48	**1.48**	1.48	1.50	1.50	**1.50**	1.50	**1.62**	**1.62**	**1.62**	**1.62**
	40	**1.57**	**1.57**	1.53	1.53	**1.53**	1.53	1.53	1.53	**1.53**	1.53	**1.68**	**1.68**	**1.68**	**1.68**
	60	**1.58**	**1.58**	1.55	1.55	**1.55**	1.55	1.56	1.56	**1.56**	1.56	**1.79**	**1.79**	**1.79**	**1.79**
80	0	1.57	1.57	1.33	1.33	1.33	1.33	1.30	1.30	1.30	1.30	1.41	1.41	1.41	1.41
	20	1.57	1.57	1.39	1.39	1.39	1.39	1.38	1.38	1.38	1.38	1.47	1.47	1.47	1.47
	40	1.57	1.57	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.46	1.53	1.53	1.53	1.53
	60	1.57	1.57	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.59	1.59	1.59	1.59
	80	1.62	1.60	1.59	1.59	1.59	1.59	1.56	1.56	1.56	1.56	1.73	1.73	1.73	1.73
100	0	**1.57**	**1.57**	1.14	1.14	**1.14**	1.14	1.14	1.14	**1.14**	1.14	**1.14**	**1.14**	**1.14**	**1.14**
	20	**1.57**	**1.57**	1.23	1.23	**1.23**	1.23	1.23	1.23	**1.23**	1.23	**1.23**	**1.23**	**1.23**	**1.23**
	40	**1.57**	**1.57**	1.31	1.31	**1.31**	1.31	1.31	1.31	**1.31**	1.31	**1.31**	**1.31**	**1.31**	**1.31**
	60	**1.57**	**1.57**	1.40	1.40	**1.40**	1.40	1.40	1.40	**1.40**	1.40	**1.40**	**1.40**	**1.40**	**1.40**
	80	**1.57**	**1.57**	1.49	1.49	**1.49**	1.49	1.49	1.49	**1.49**	1.49	**1.49**	**1.49**	**1.49**	**1.49**
	100	**1.57**	**1.57**	1.57	1.57	**1.57**	1.57	1.57	1.57	**1.57**	1.57	**1.57**	**1.57**	**1.57**	**1.57**

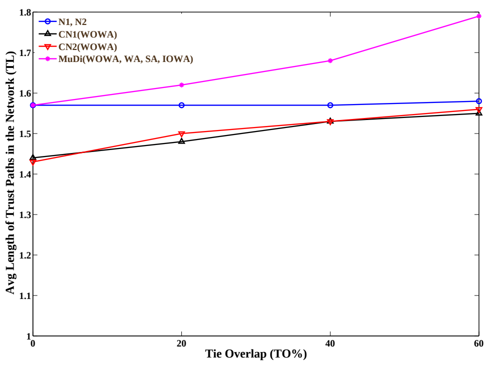
The TL metric for CN1 and CN2 (for shortest path recommendation) in Table 5.7 and Figures 5.6(c) and 5.6(d) also shows an improvement, with an increase in PO compared with N1 and N2, as its value for all the aggregation techniques (S, WA, WOWA, IOWA) drops to (1.52, 1.50) for (CN1, CN2) at $[40\%PO, 0\%TO]$ and (1.14, 1.14) for (CN1, CN2) at $[100\%PO, 0\%TO]$, which were (1.57, 1.57) in N1 and N2 respectively. The reason for having the same TL value for all the aggregation techniques is that TL now depends on the number of hops between N_p, unlike the strongest path recommendation in which it depends on the strength of the path.



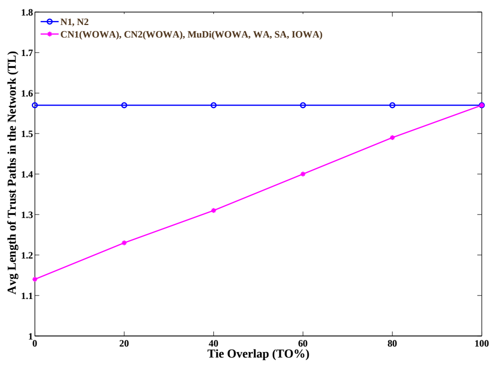
(a) Participant Overlap(PO) = 60%



(b) Participant Overlap(PO) = 100%



(c) Participant Overlap(PO) = 60%



(d) Participant Overlap(PO) = 100%

Figure 5.6: Sample values of TL selected from Table 5.4, 5.6 for depiction. N1 and N2 represent original networks, MuDi(S), MuDi(WA), MuDi(WOWA) and MuDi(WOWA) represent TL in MuDi for S , WA, WOWA and IOWA trust aggregation techniques. CN1 (WOWA) and CN2 (WOWA) show TL metric from sub-networks CN1 and CN2 in MuDi for WOWA aggregation technique. Figures 5.6(a) and 5.6(b) are from strongest path algorithm while Figures 5.6(c) and 5.6(d) are from shortest path algorithm.

The TL in MuDi (for strongest path recommendation) behaves similarly to that in CN1 and CN2 (for strongest path recommendation) and its value for all the aggregation techniques (S, WA, WOWA, IOWA) decreases with an increase in PO. The worst performance among all the techniques lies with S as its value becomes greater than both the N1 and N2 i.e. 3.04 compared to (2.68, 2.41) in (N1, N2) at $[40\%PO, 0\%TO]$. WA and IOWA metrics are best with TL less than N1 and N2 for all values of PO and TO. For WOWA, although TL decreases with an increase in PO, its value becomes greater than either of the networks N1 and N2 at some data points with low PO. However, if coupled with the TS metric for same values of PO and TO, it appears to be the best approach among the aggregation techniques due to its trust preservance feature. The noise and non-uniformity in the TL metric is due to its dependence on TS as, to achieve the maximum TS, the trust algorithm can even select longer trust paths.

The TL metric in MuDi (for shortest path recommendation) is also the same for all the aggregation techniques due to considering the length of path rather than the strength of path. Its value only decreases for $PO \geq 80$, but then increases with an increase in TO and at $[80\%PO, 80\%TO]$ it becomes 1.73, that is, even greater than the original networks N1 and N2. This is due to the decreasing number of non-overlapping ties with an increase in TO which results in fewer new trust paths. For $[40\%PO, 0\%TO]$, TL is higher (i.e. 1.69) than (N1, N2) (1.57, 1.57), which becomes even higher (i.e 1.79) at $[40\%PO, 30\%TO]$. The maximum number of new shortest paths is generated at $[100\%PO, 0\%TO]$ and as a result TL drops to 1.14 corresponding to (1.57, 1.57) in (N1, N2) but at [100%PO, 100%TO] it again becomes 1.57 (equal to N1 and N2) after 100% overlap between the networks.

To analyse the behaviour of TS and TL metrics over varying percentages of PO and TO, the next section presents an in-depth analysis of these metrics with respect to the hypothesis claim.

5.4.2 Trust Data Analysis

The aim of our experiment was to aggregate trust information from MuDi social networks without affecting the integrity of that information. We can define this as preserving the trust values from the original networks (N1, N2) in the sub-networks (CN1, CN2) of the consolidated networks.

From analysing the TS metric presented, it can be seen that WOWA technique better aggregates the trust from MuDi social networks, as it respects the integrity of trust in N1 and N2. S simply amplifies the trust and WA naively dampens down the trust, while IOWA escalates trust from one of the trust sources but, as WOWA fuses trust data available on the ties bearing in mind their importance, it gives a more balanced aggregation (distinguishing the *absence of trust* from *distrust*).

Referring to our hypothesis, we can say that a consolidation approach better preserves the trust values if the trust values in the original networks are similar to those in the relevant sub-networks of the consolidated network. We expect consolidation to create some differences, but that each trust relationship would be as likely to rise as to fall, and therefore when averaged across all ties it remains approximately stable. We can check this by comparing TS between the networks for each consolidation technique.

We expect the second part of our hypothesis, the opening up of many additional trust paths, to manifest through the TL metric that measures the average length of trust paths. The data show that TL in the MuDi network is dependent on the Participant Overlap PO. When PO is low it creates a path bottleneck in the consolidated network, and TL is higher than for the original networks, but when PO is high the increase in connections causes TL to fall. Additionally, it can be seen that TL in each of the sub-networks CN1 and CN2 is lower than in the corresponding original network N1 and N2 respectively, regardless of the value of PO. This shows that the additional trust paths are being created.

Our numerical simulation showed that the WOWA consolidation of MuDi social networks is a productive approach that preserves the integrity of the trust values (as measured by an increase in TS, average strength of ties) while creating new trust paths (as measured by decrease in TL, average length of trust paths). At low PO it creates an opportunity for users to know and interact with many new users who are not part of their original networks, and hence creates ties between people from networks of different backgrounds. On the other hand, at high PO and TO WOWA consolidation helps in refining trust values by combining different perspectives of trust between participants in different networks.

To test whether this apparent preservence of trust integrity by WOWA unlike other techniques is statistically significant, TS and TL metrics from WOWA were tested for statistical significance with WA and IOWA. The statistical significance test for this analysis is a two-tailed paired T-Test that evaluates whether the apparently more stable WOWA aggregation metric generates significantly better results than WA and IOWA or happens just by chance. This test will generate a p-value and, if the following hypothesis is true, means that it is true that WOWA is significantly better than both WA and IOWA.

Hypothesis: If $p \leq 0.05$, it means that TS and TL metrics from WOWA are significantly better in preserving integrity of trust from N1 and N2 than WA and IOWA.

Null Hypothesis: If $p > 0.05$, it means that TS and TL metrics from WOWA are not significantly better in preserving integrity of trust from N1 and N2 than WA and IOWA.

5.4.2.1 Statistical significance of WOWA over WA for average strength of trust ties (TS) and average length of trust paths (TL)

Table 5.8 shows the p-value for two types of *PO* related to TS. The, first four measurements represent the p-value for varying *PO*, while the last measurement, that is, *overall* represents the collective performance of the system including data for all values of *PO*.

Looking at the p-value, it can be seen that for either value of *PO* and for both strongest and shortest path recommendation mechanisms $p \leq 0.05$, which proves our hypothesis that WOWA is significantly better than WA.

Table 5.8: Statistical Significance (p-value) between corresponding TS metrics for WA and WOWA from CN1, CN2 and MuDi.

Networks / *PO*	Strongest Path Algo			Shortest Path Algo		
	CN1	CN2	MuDi	CN1	CN2	MuDi
40%	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
60%	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
80%	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
100%	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Overall	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Analysing the same p-value in Table 5.9 for TL metric however shows somewhat different results. Here results from WA and WOWA are not significantly different apart from $PO = 60\%$ or when *overall* performance is measured. That is, $p \leq 0.05$ only holds true for $PO = 60\%$ and for *overall* performance of the system. The p-value for the shortest path recommendation mechanism cannot be evaluated for TL metric as considering path length with respect to the number of ties generates exactly same values for both WA and WOWA.

Table 5.9: Statistical Significance (p-value) between corresponding TL metrics for WA and WOWA from CN1, CN2 and MuDi.

Networks / *PO*	Strongest Path Algo			Shortest Path Algo		
	CN1	CN2	MuDi	CN1	CN2	MuDi
40%	0.09	0.05	0.05	-	-	-
60%	0.02	0.03	< 0.01	-	-	-
80%	0.10	0.30	0.06	-	-	-
100%	0.06	0.06	0.06	-	-	-
Overall	< 0.01	< 0.01	< 0.01	-	-	-

5.4.2.2 Statistical significance of WOWA over IOWA for average strength of trust ties (TS) and average length of trust paths (TL)

Table 5.10 shows the p-value for statistical significance of WOWA over IOWA for TS metric and the results show that, apart from $PO = 40\%$ (for strongest path recommendation), p $\leq$ 0.05. This means that the WOWA is significantly better than IOWA for CN1, CN2 and MuDi. p-value for CN1 in both the strongest and shortest path algorithms is not as small as for CN2 and MuDi, and the reason is the high importance of the trust values from N1 (unlike N2), which generates IOWA-aggregated metrics for N1 more similar to those generated by WOWA aggregation than N2.

Table 5.10: Statistical Significance (p-value) between corresponding TS metrics for IOWA and WOWA from CN1, CN2 and MuDi.

Networks / *PO*	Strongest Path Algo			Shortest Path Algo		
	CN1	CN2	MuDi	CN1	CN2	MuDi
40%	0.08	< 0.01	< 0.01	0.05	< 0.01	< 0.01
60%	0.05	< 0.01	< 0.01	0.04	< 0.01	< 0.01
80%	0.03	< 0.01	< 0.01	0.02	< 0.01	< 0.01
100%	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Overall	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

The p-value for TL (Table 5.11) is somewhat different from the TS metric and it disproves the hypothesis for all values of PO for both CN1 and CN2. However, for MuDi version of the networks, it proves our hypothesis for all values of PO apart from $PO = 100\%$. Similar to the case of WA, here the shortest paths for WOWA and IOWA are exactly same and hence the p-value for such a data cannot be evaluated and are left blank.

Table 5.11: Statistical Significance (p-value) between corresponding TL metrics for IOWA and WOWA from CN1, CN2 and MuDi.

Networks / *PO*	Strongest Path Algo			Shortest Path Algo		
	CN1	CN2	MuDi	CN1	CN2	MuDi
40%	0.97	0.83	< 0.01	-	-	-
60%	0.50	0.78	0.01	-	-	-
80%	0.55	0.55	0.02	-	-	-
100%	0.69	0.69	0.69	-	-	-
Overall	0.97	0.59	< 0.01	-	-	-

5.4.3 Discussion

The increasing use of multiple heterogeneous social networks, both explicit and implicit, offers an opportunity to refine trust calculations by consolidating multiple trust networks into a single network for analysis. However, consolidating trust networks is non-trivial due to variance in node and tie overlap, differences in the importance of networks, and differences in expressing trust.

In this experiment we have presented a numerical simulation of what happens when different trust networks (with the characteristics of real-world networks) are consolidated using one of the four strategies: S, WA, WOWA and IOWA. In our experiment we varied participant and tie overlap, and recorded the effect on average strength of ties and average length of trust paths for the whole consolidated network (MuDi), and the sub-networks (CN1, CN2) representing the original networks (N1, N2).

Our analysis revealed that the summation (S) strategy results in an inflation of trust values, while the weighted average (WA) results in dampened trust values. Induced Ordered Weighted Averaging (IOWA) preserved the integrity of trust in one of the networks, but severely distorted for the other. However, the Weighted Ordered Weighted Averaging (WOWA) strategy has much improved performance, in that it better preserved the integrity of the trust compared to WA ($p < 0.01$) and IOWA ($p < 0.01$), while also being better than WA at creating shorter trust paths ($p < 0.01$).

Our experiment showed that WOWA can be used to consolidate trust networks without damaging trust values. However, it is still not clear whether the changes to trust values caused by consolidation actually increase their *quality* in terms of their similarity to the trust actually felt by those individuals.

To test this, our next step was to attempt this consolidation activity with two real social networks, looking at professional and co-authorship networks, and then to perform a quantitative evaluation with actual users via a survey to compare actual trust values with those in the original and consolidated networks.

We have shown that a WOWA consolidation strategy can effectively combine multiple trust networks, providing evidence that trust values derived from MuDi social networks can be merged to create new trust paths without damaging trust values. Our hope is that this approach can be used to create more reliable trust calculations that take advantage of our increasingly rich and varied online interactions.

5.5 Conclusions

This chapter ran a simulation experiment and evaluated the idea of consolidating multiple social networks using different data fusion techniques. The simulation was run for different values of PO and TO and TS metric showed that the WOWA generates more stable aggregates of trust over all values of PO and TO, thus improving the overall quality of trust. Furthermore, TL also showed improved results due to emergence of more trust paths over high values of PO.

Results showed that naive methods of consolidation damage trust, if trust values available from all the networks are not considered alongside the importance of the networks themselves, and vice versa. The statistical significance level indicates that the analysed response of the proposed consolidation strategy is not happening by chance, rather that the behaviour of WOWA is consistently better and works for different values of PO and TO.

In the next chapter, the proposed WOWA trust aggregation technique is run using real-world data from two professional social networks to see whether the WOWA technique actually improves trust metrics, compared to using only those from individual networks, in terms of users' actual trust perceptions.

Chapter 6

Experiment II - Real World Data Analysis of MuDi Trust Aggregation

6.1 Introduction

This chapter examines the proposed idea of consolidating MuDi social networks by analysing the accuracy of aggregated trust metrics from two professional networks and compares it with those collected from actual users using an interpersonal trust survey. It incorporates two MuDi social networks and consolidates them to generate an updated trust metrics between the participants using the semantic web framework proposed in Chapter 3.

The recommendation of the simulation experiment about WOWA being the better aggregation approach was used for aggregating multiple trust values. Direct trust information between users was preferred, if available, otherwise it was evaluated for distant nodes using both the strongest and shortest trust paths to see which better matches the actual trust provided by users. A survey was conducted that asked a set of professional trust questions from a group of people working in a university research group about other researchers working in the same environment using an online questionnaire specifically designed for the purpose. Finally, the similarity of the trust values calculated using a single or consolidated MuDi networks was compared to those gathered by the survey, in order to show whether trust values generated by consolidated MuDi networks are more similar to perceived trust than trust values generated by a single component network.

6.2 The Real World Social Networks

A pair of professional networks (co-authorship and collaboration networks) managed by the University of Southampton was selected to represent proxy trust between a pair of individuals. The co-authorship network was extracted from ePrints' Soton domain and includes researchers who publish articles. Eprints[4] is an open publication archive that hosts research articles of people working at University of Southampton. For this work, publications of those working in the WAIS (Web and Internet Science) group are selected. WAIS projects[5] network contains details about the funded projects researchers from the University of Southampton are involved in. Data from both active and past projects is considered for this experiment.

The frequency of co-authorship (evaluated using SPARQL SELECT QUERY 12) was considered as a proxy for trust, while the collaboration network was harvested from the WAIS projects domain that represents pairs of researchers working on the same projects, with frequency of collaboration (calculated using SPARQL SELECT QUERY 13) attached as a proxy trust value.

SPARQL SELECT Query 12 Calculating Co-author Frequency

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX ep: <http://eprints.org/ontology/>

SELECT (COUNT(?epuri1) AS ?coauthorfreq)
FROM NAMED <http://eprints.soton.ac.uk>
WHERE {
ep:EPrint rdf:type rdfs:Class.
GRAPH <http://eprints.soton.ac.uk> {
        ?epuri1 rdf:type ep:EPrint.
        ?epuri2 rdf:type ep:EPrint.
        ?epuri1 dct:creator <"""+epcreatoruri1+""">.
        ?epuri2 dct:creator <"""+epcreatoruri2+""">.
        FILTER (?epuri1 = ?epuri2). }
}
```

To represent the implicit publication information between authors, ePrints uses the Dublin Core vocabulary (Board, 2012) (*dct:creator* for specifying authors of the paper), while WAIS project membership uses an ECS (Electronics and Computer Science) ontology (ECS, 2013) with property *ecs:memberOf* presenting the list of members working on a certain project. The SPARQL query was run over these datasets (in Sesame) to find co-author/collaboration frequency (proxy trust) between pairs of researchers. The Sesame repository stores RDF data from ePrints, WAIS and the consolidated version of these networks as named graphs with the *context* parameter (in sesame, a triple can be stored as a *quad*, and the additional parameter *context* represents the provenance of the triple - see Section 2.6.2) acting as the name for each graph. The OWL-Lite plugin was used to let us store our ontology and inferred triples from RDF data.

SPARQL SELECT Query 13 Calculating Collaboration Frequency

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ep: <http://eprints.org/ontology/>
PREFIX ecs: <http://rdf.ecs.soton.ac.uk/ontology/ecs#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT (COUNT(?epuri1) AS ?collaboratefreq)
FROM NAMED <http://wais.soton.ac.uk/projects>
WHERE {
foaf:Project rdf:type rdfs:Class.
GRAPH <http://wais.soton.ac.uk/projects> {
        ?project1 rdf:type foaf:Project.
        ?project2 rdf:type foaf:Project.
        <”””+epcreatoruri1+”””> ecs:memberOf ?project1.
        <”””+epcreatoruri2+”””> ecs:memberOf ?project2.
        FILTER (?project1 = ?project2). }
}
```

Both of these networks are in the same environment, so there is a significant *PO* (Participant Overlap) and *TO* (Tie Overlap), with the collaboration network nearly a subset of the co-authorship network, discounting users outside the university who work on projects. The *PO* and *TO* with respect to the WAIS were 51% and 78% while for the Eprints, they were 2% and 1.4% respectively. Table 6.1 shows description of the network parameters used in this experiment. Both these networks contain bidirectional symmetric trust, as co-authorship and collaboration represents the same trust values in both directions.

Table 6.1: Network and consolidation parameters used for the real-world experiment for measuring the accuracy of aggregated trust.

Parameters	ePrints	WAIS
N	3286	154
PO	2%	51%
TO	1.4%	78%
p_{WOWA}	[0.8,0.2]	
w_{WOWA}	[1,0.5]	

6.2.1 Resolving Multiple Co-referred Identities

Both the selected networks are from the same domain and the majority of participants work in the same environment, but they contain different URIs for participants due to different namespaces used in each of the networks. So a logical or label comparison needed to be undertaken to find co-referring URIs, since a simple solution of URI comparison would not work here - see Section 2.5.2.1. Logical comparison identified co-referred users by matching Inverse Functional Properties (IFPs) of individuals from multiple networks; those with the same values of IFPs were termed co-referred users. In the selected networks there is no single IFP, so this mechanism would not work.

Label comparison was conducted using owl Datatype Properties (*owl:DPs*) of participants (for example *foaf:familyName* and *foaf:givenName*) in both the selected networks. The ePrints network contains the target *owl:DPs*, but the collaboration network does not include any such information. The URIs used in the collaboration network were, however, linked with AKT (Advanced Knowledge Technologies) project[1] local URIs (also developed at the University of Southampton) using *owl:sameAs* predicate, and this project also includes users' data properties (*foaf:familyName* and *foaf:givenName*). Using the linked data, label comparison was conducted with *owl:DPs* from AKT project local URI corresponding to the equivalent collaboration network URI (SPARQL SELECT Query 14 snippet between lines 10 and 11 performs this function). FILTER statement at line 11 in SPARQL SELECT Query 14 classified co-referred users from both the networks. Those users not included in the co-referred category were classified as non-co-referred users.

SPARQL SELECT Query 14 Coreference Resolution

```
 1: PREFIX foaf: <http://xmlns.com/foaf/0.1/>
 2: PREFIX akt: <http://www.aktors.org/ontology/portal#>
 3: PREFIX owl: <http://www.w3.org/2002/07/owl#>
 4:
 5: SELECT DISTINCT ?eprintsuri ?waisuri
 6: FROM NAMED <http://eprints.soton.ac.uk>
 7: FROM NAMED <http://wais.soton.ac.uk/projects>
 8: WHERE {
 9: GRAPH <http://eprints.soton.ac.uk> {
            ?eprintsuri a foaf:Person.
            ?eprintsuri foaf:givenName ?eprintsgn.
            ?eprintsuri foaf:familyName ?eprintsfn. }
10: GRAPH <http://wais.soton.ac.uk/projects> {
            ?waisuri a foaf:Person.
            ?akturi owl:sameAs ?waisuri.
            ?akturi akt:given-name ?waisgn.
            ?akturi akt:family-name ?waisfn. }
11: FILTER (?eprintsgn=?waisgn && ?eprintsfn=?waisfn)
12: }
```

After having co-referred URIs of overlapping users from both the networks, the consolidated graph needed to be annotated with a new single URI (i.e. *mudiURI*) corresponding to the co-referred pair for making trust annotations. For linking it to the

[1] http://www.aktors.org/akt/

individual networks, set of triples having *owl:sameAs* relationship were added, linking *mudiURI* in consolidated graph to each of the co-referred URI in the corresponding individual network. This was performed using SPARQL INSERT Query 15 over the consolidated named graph by mentioning the name of the consolidated graph before each triple. *CoreferedURIePrints* and *CoreferedURIWAIS* represented URIs in each of the multiple distributed networks and *mudiURI* was the new URI in the consolidated graph. Although *owl:sameAs* is a symmetric property, here triples were explicitly added in both directions because Sesame classifies inferred triples in a default graph rather than in the same graph, and it would have been impossible for it to have been machine read if it were not explicitly added to the graph.

SPARQL INSERT Query 15 Coreference (owl:sameAs) Annotations

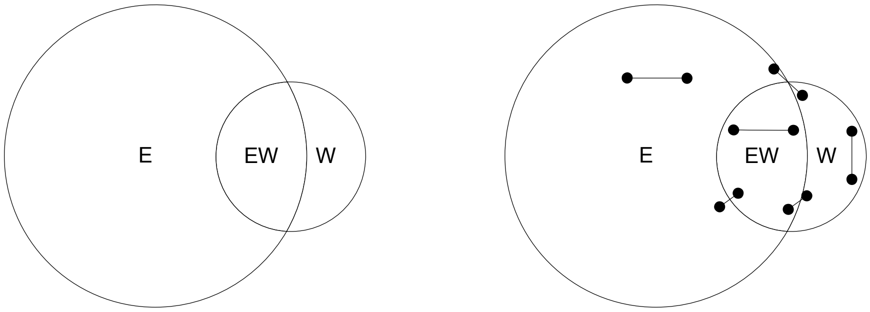
```
PREFIX owl: <http://www.w3.org/2002/07/owl#>

Insert DATA {
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+coreferredURIePrints+”””> owl:sameAs <”””+mudiURI+”””>.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+coreferredURIWAIS+”””> owl:sameAs <”””+mudiURI+”””>.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+mudiURI+”””> owl:sameAs <”””+coreferredURIePrints+”””>.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+mudiURI+”””> owl:sameAs <”””+coreferredURIWAIS+”””>.}
}
```

6.2.2 Multiple Distributed Trust Consolidation

This section identifies Trust Aggregation Scenarios (TASs) for real-world networks using the procedure described in Section 4.4. For the selected pair of networks, the three TASs resulted in six different types of participant pairs (shown in Figure 6.1(b)) from three different regions (E, EW, W) of the Venn diagram (Figure 6.1(a)) using the Equation 4.7:

$$\begin{aligned} TASs &= PC_{FRs} + PC_{CRs} \\ TASs &= 3 + 3 \\ &= 6 \end{aligned} \tag{6.1}$$

Of these TASs, one type of participant pairs (N_p) belonged to the overlapping region (EW), having complete trust information between users, two types of participant pairs belonged to non-overlapping regions (E, W) and three belonged across different regions (E→ EW, W→ EW, E→ W), resulting in partial trust information. This generates the two types of trust metrics, based on the availability of trust information - already discussed in cases 1 and 2 in Section 4.5. Two SPARQL queries were written for mining information corresponding to these two cases.



(a) Venn digram showing overlapping and non-overlapping potions of ePrints (co-authorship) and WAIS projects (collaboration) professional trust networks

(b) Detailed venn diagram with all possible pair of participants belonging to different regions in the set of networks

Figure 6.1: Two sample venn diagrams depicting different regions (E for ePrints, W for WAIS and EW for overlapping portion) and the 6 types of pairs that can exist within those regions.

6.2.2.1 Overlapping Trust Aggregation

Overlapping trust aggregation represents a scenario where both users of the pair are in both the networks and the region EW in Figure 6.1 contains such pairs of users. Trust information from co-referred pairs was available in both the networks that needed to be aggregated.

SPARQL SELECT Query 16 implemented the overlapping trust aggregation scenario and showed three graphs from the triplestore with the first two graphs, *<http://eprints.-soton.ac.uk>* and *<http://wais.soton.ac.uk/projects>*, representing ePrints and WAIS professional networks respectively and the third graph, *<http://consolidatedmudinetworks-.com>*, the consolidated version of ePrints and WAIS networks. The consolidated graph was already annotated for co-referred URIs in Section 6.2.1 and it contained the *owl:sameAs* property with MuDi URIs of users having links with both the co-referred URIs in the existing individual networks.

If the target pair of users under consideration were co-referred users, meaning that their *owl:sameAs* links existed in the consolidated graph, (represented in query as co-referred trustor, with variables *trustoreprints, trustorwais*, and the co-referred trustee with variables, *trusteeeprints, trusteewais*), then the trust values between them from both networks ePrints and WAIS, *?pvalueeprints* and *?pvaluewais*, along with their MuDi URIs *?muditrustorURI* and *?muditrusteeURI* were returned for the WOWA aggregation function to generate a single trust value.

SPARQL SELECT Query 16 Overlapping Trust Aggregation

```
PREFIX trust: <http://trustontology.com/owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?muditrustorURI ?muditrusteeURI ?pvalueeprints ?pvalue-
wais
FROM NAMED <http://eprints.soton.ac.uk>
FROM NAMED <http://wais.soton.ac.uk/projects>
FROM NAMED <http://consolidatedmudinetworks.com>
WHERE {
GRAPH <http://eprints.soton.ac.uk> {
        ?trustrelep a trust:TrustRelationship.
        ?trustrelep trust:trustor <”””+trustoreprints+”””>.
        ?trustrelep trust:trustee ?trusteeeprints.
        ?trustrelep trust:processedvalue ?pvalueeprints. }
GRAPH <http://wais.soton.ac.uk/projects> {
        ?trustrelws a trust:TrustRelationship.
        ?trustrelws trust:trustor <”””+trustorwais+”””>.
        ?trustrelws trust:trustee ?trusteewais.
        ?trustrelws trust:processedvalue ?pvaluewais. }
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+trustoreprints+”””> owl:sameAs ?muditrustorURI.
        <”””+trustorwais+”””> owl:sameAs ?muditrustorURI.
        ?trusteeeprints owl:sameAs ?muditrusteeURI.
        ?trusteewais owl:sameAs ?muditrusteeURI. }
}
```

6.2.2.2 Singular Trust Re-evaluation

In trust aggregation, there are situations when one of the users in the pair is present in one of the network while the other user is either in the same network as the first user or is a co-referred user in a part of multiple networks. This is represented by a non-overlapping region or cross-regions/single network scenarios and Figure 6.1 includes user pairs, for example, E $\rightarrow$ E, E $\rightarrow$ EW, W $\rightarrow$ EW and so on. The unique aspect about all such pairs is the availability of singular trust information that needs to be re-evaluated.

SPARQL SELECT Query 17 modelled this trust aggregation scenario for a sample trustor (represented using variable *trustoreprints*) to be in the region E, shown in query as the member of graph *<http://eprints.soton.ac.uk>*. The trustee (represented using *?trustee*) either belonged to the same region as the trustor, effectively to the same graph in the triplestore or, if the *owl:sameAs* property exists for it in the consolidated graph, then in the cross-region. The additional OPTIONAL clause solved this condition and its result determined the location of the *?trustee*. If the result of this clause was '*false*', it meant *?trustee* was in the same region E as of *trustoreprints* and hence *?trustee* URI and its trust value (represented in query 17 with the variable *?pvalueeprints*) with the trustor was returned as an ouput of this query. If the result of this clause was '*true*', it meant *?trustee* existed in the cross-regions EW and the MuDi URI (*?mudiuri*) was returned along with the trust value (*?pvalueeprints*) between them from the ePrints network. The trust value between user pairs in both the cases is a single value extracted from ePrints network because, although one of the users in the pair may be a co-referred user, the other always exists in only one of the network.

SPARQL SELECT Query 17 Cross-Networks Trust Re-evaluation

```
PREFIX trust: <http://trustontology.com/owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?trustee ?pvalueeprints ?mudiuri
FROM NAMED <http://eprints.soton.ac.uk>
FROM NAMED <http://consolidatedmudinetworks.com>
WHERE {
GRAPH <http://eprints.soton.ac.uk> {
        ?trustrelep a trust:TrustRelationship.
        ?trustrelep trust:trustor <"""+trustoreprints+""">.
        ?trustrelep trust:trustee ?trustee.
        ?trustrelep trust:processedvalue ?pvalueeprints. }
OPTIONAL {
GRAPH <http://consolidatedmudinetworks.com> {
        ?trustee owl:sameAs ?mudiuri. }
}
}
```

6.2.3 Annotating Updated Trust

The consolidated graph was annotated with the updated trust between participants, once aggregated and re-evaluated trust metrics were calculated between them from the MuDi networks. It used the trust ontology proposed in Section 3.6 and added all the annotations in the consolidated graph to the updated trust values. Direct trust estimations use the WOWA data fusion technique for aggregating multiple trust metrics, while trust for distant participants was evaluated using the decay-based trust inference mechanisms discussed in Section 4.7.

SPARQL INSERT Query 18 added trust data to the consolidated graph *<http://-consolidatedmudinetworks.com>* with all the aggregated trust measurements.

SPARQL INSERT Query 18 Annotating Trust Annotations

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX trust: <http://trustontology.com/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

Insert DATA {
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> a trust:TrustRelationship.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:trustor <”””+mtrustoruri+”””>.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:trustee <”””+mtrusteeuri+”””>.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:scope ’””+tSubject+”””’^^xsd:string.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:fuzzyvalue ’””+fValue+”””’^^xsd:string.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:processedvalue ’””+str(pValue)+”””’^^xsd:float.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:absolutevalue ’””+str(aValue)+”””’^^xsd:integer.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:type ’””+ttype+”””’^^xsd:string.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:pathlength ’””+plength+”””’^^xsd:integer.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:process ’””+tprocess+”””’^^xsd:string.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:metric ’””+tmetric+”””’^^xsd:string.}
GRAPH <http://consolidatedmudinetworks.com> {
        <”””+uri+”””> trust:aggregationtechnique ’””+tat+”””’^^xsd:string.}
}
```

6.3 Experiment Design

To analyse the aggregated trust measurements from consolidated pair of networks for accuracy, a survey experiment was designed to collect professional proxy trust between researchers in the networks. It constituted two parts. The first portion comprised trust-related questions while the next section asked an expert recommendation question.

6.3.1 Trust Survey

To judge the accuracy of calculated trust between participants after consolidating MuDi trust networks, the designed survey application collected reports of the actual professional trust between participants in real life. This was a web application that presented each participating user with a set of related people from one or both of the networks, based on the presence of the user in the networks, and asked a set of proxy trust questions (which represented implicit trust in the professional context), that helped us to analyse trust between them. The data layer of the application was a Sesame triplestore and contained the trust-annotated ePrints, the WAIS projects individual social networks, and their consolidated version as separate named graphs. As the WAIS project's dataset did not carry meta-data information about users, this was added as an RDF graph into projects data from AKT project[1] developed at the University of Southampton. The SPARQL endpoint of the triplestore provided an interface to exploit RDF data by writing SPARQL queries.

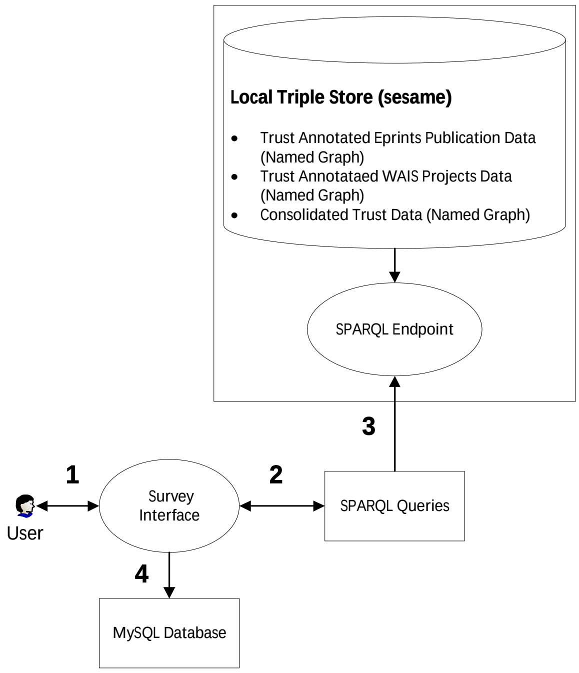


Figure 6.2: Interaction diagram representing different components of the survey application

Figure 6.2 shows the interaction diagram of a *user* with the survey application.

1. *User* enters the required information to Log-in to the survey interface.

2. The application sends that information to the SPARQL endpoint of the Sesame triplestore by generating SPARQL queries at the runtime.

3. The result of the queries returns the set of colleagues to be evaluated for proxy trust and are displayed to the logged-in *user*.

4. The *user* answers the proxy trust questions and submits the survey back to the server, and results are stored in MySQL database.

Further details about the survey and the specific questions asked to each of the partiticpating user are discussed in Sections 6.3.1.1, 6.3.1.2, 6.3.1.3 and 6.3.2.

6.3.1.1 Participant Selection

There are two aspects of participant selection in this survey. The first set of people comprised those participating (*rating* participants) in the survey and the second set was a selected group of people (*rated* participants) about whom *rating* participants expressed their implied trust.

The selection of the *rating* participants was uniformly probabilistic within the ePrints and WAIS projects domain as anyone in the dataset had an equal opportunity to become part of the survey but needed to log-in first (using *firstname* and *lastname*), while the selection of the *rated* participants used information from the dedicated ego-centric network extracted with logged-in user acting as an ego node. As there was a decay of trust along the trust path in the simulated model of trust, the trust measurements for indirectly connected participants also needed to be recorded to judge the significance of this claim. Keeping this in mind, *rated* participants were selected as allegedly belonging to path lengths of *one*, *two* and *three* by the *rating* participant.

The number of the *rated* participants also mattered for the *rating* participants in the survey. To get a good range of data points, we needed to ask as many questions as possible, whilst keeping the survey to a manageable size. Through a *focus group* session (questionnaire in Appendix D), a figure of *eight* people emerged as an appropriate compromise. The set of *rated* participants belonged to the neighbourhood of *three* steps, with *three* people each from Path length *one* and *two* and *two* people from Path length *three*. If the *rating* participant was present in both the networks, then *four* of the participants were selected from each of the networks, otherwise all the *rated* participants were selected from one of the networks.

6.3.1.2 Questionnaire

As we were to measure the quality of consolidated trust in a pair of MuDi professional social networks, this included co-authorship and collaboration networks, and this survey aimed to extract the proxy trust that participants infer about each other in a professional context. Each of the rating participants was presented with a set of questions to help us collect implied trust data about other network members and a list of experts relating to one of their research area. These measurements were compared with system readings to evaluate the performance of trust metrics generated by the system. There are techniques for collecting human observations through analysing the performance of the proposed systems in the literature, for example Zhang et al. (2008) asked a faculty member and two students to mention the number of their publications along with other research profile attributes to assess the precision of an expert-finding algorithm. Similarly, Sim and Crowder (2004) interviewed a set of nine people from an engineering design department using sample questions from their fields to evaluate the performance of the designed expert recommendation techniques.

The substance of the survey, and especially selection of the questions is also tricky in studies involving human participation. An assurance of data privacy and ethics while asking such questions increases the challenge in the context of trust, as it aims to obtain one's personal sentiments about others. In this case, proxy trust evaluation survey also tried to extract such information and modelled it using two questions. *First*, *rating* participants were asked about their *past work* experience with the person, and, *second*, the likelihood of them *working together in future* should there be the opportunity. Participants could select one out of the five options in both the questions and their selection would be stored in the database as integer values in the range [1 to 5] (corresponding to left to right in the survey in increments of 1 - see Figure 6.4)). Numeric data from this portion of the survey were mormalised, by dividing all values with the maximum among them (i.e. it can be 5 at most), and then comparing with the data available from the system, which is already in the range [0,1].

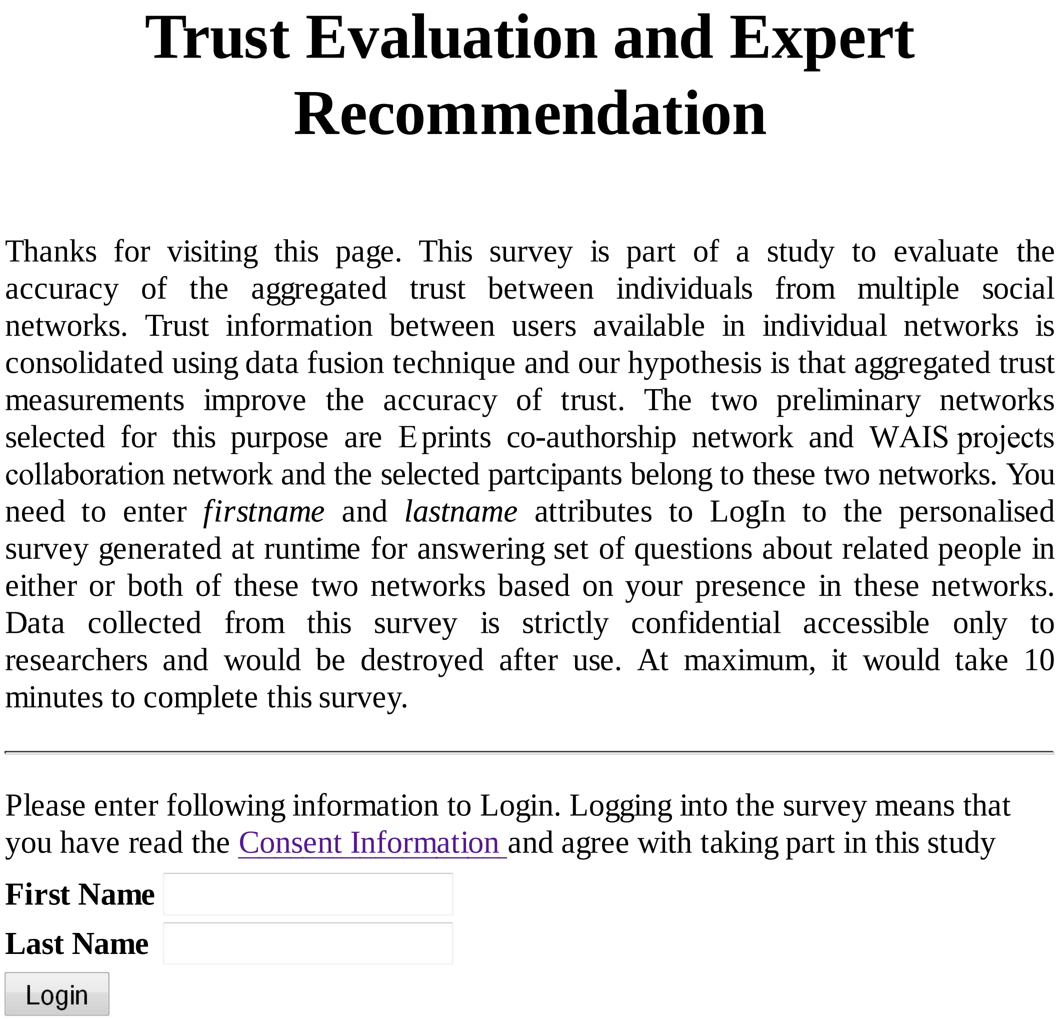
Looking further into the academic and professional networks, links between users may be divided into multiple categories depending upon their roles in these networks, for example, supervisor, colleague and so on. This gave us the opportunity to analyse trust in each of these categories, particularly. The *third* question in the survey served that purpose by asking *rating* participant briefly to explain their relationship with each *rated* person separately. A set of relationships provided by *rating* participants were then classified to generate relationship categories. Comparative analysis held for all type of relationships generically in the above hypothesis was now repeated for people from each of these categories specifically to analyse the system measurements regarding which category was best improved compared with the other relationship categories.

To obtain the consent of users about ethical standards and perception about the accuracy of questions, a prior *focus group* study had assessed the quality of the questions and, based on the recommendations of the members, the set of questions shown in Figure 6.4 was finalised for the survey.

Questions were categorised into two set. The *first* question asked *rating* participants about their familiarity with the *rated* person. Only if the *rating* person says '*Yes*' to the *first* question was the *second* set presented, otherwise the *rating* participant could move on to the next person in the survey. This is because *rated* participants were at varied path lengths from a *rating* participant and it might be possible that the person participating in the survey not to be at all familiar with that person.

6.3.1.3 Application Interface

The web-based application to collect real life trust and expert recommendation values from participants was designed in the Django [2] web Framework for Python. It included three web pages that complete the process of extracting trust measurements from participants belonging to the ePrints and WAIS projects professional social networks.



Trust Evaluation and Expert Recommendation

Thanks for visiting this page. This survey is part of a study to evaluate the accuracy of the aggregated trust between individuals from multiple social networks. Trust information between users available in individual networks is consolidated using data fusion technique and our hypothesis is that aggregated trust measurements improve the accuracy of trust. The two preliminary networks selected for this purpose are Eprints co-authorship network and WAIS projects collaboration network and the selected partcipants belong to these two networks. You need to enter *firstname* and *lastname* attributes to LogIn to the personalised survey generated at runtime for answering set of questions about related people in either or both of these two networks based on your presence in these networks. Data collected from this survey is strictly confidential accessible only to researchers and would be destroyed after use. At maximum, it would take 10 minutes to complete this survey.

Please enter following information to Login. Logging into the survey means that you have read the Consent Information and agree with taking part in this study

First Name

Last Name

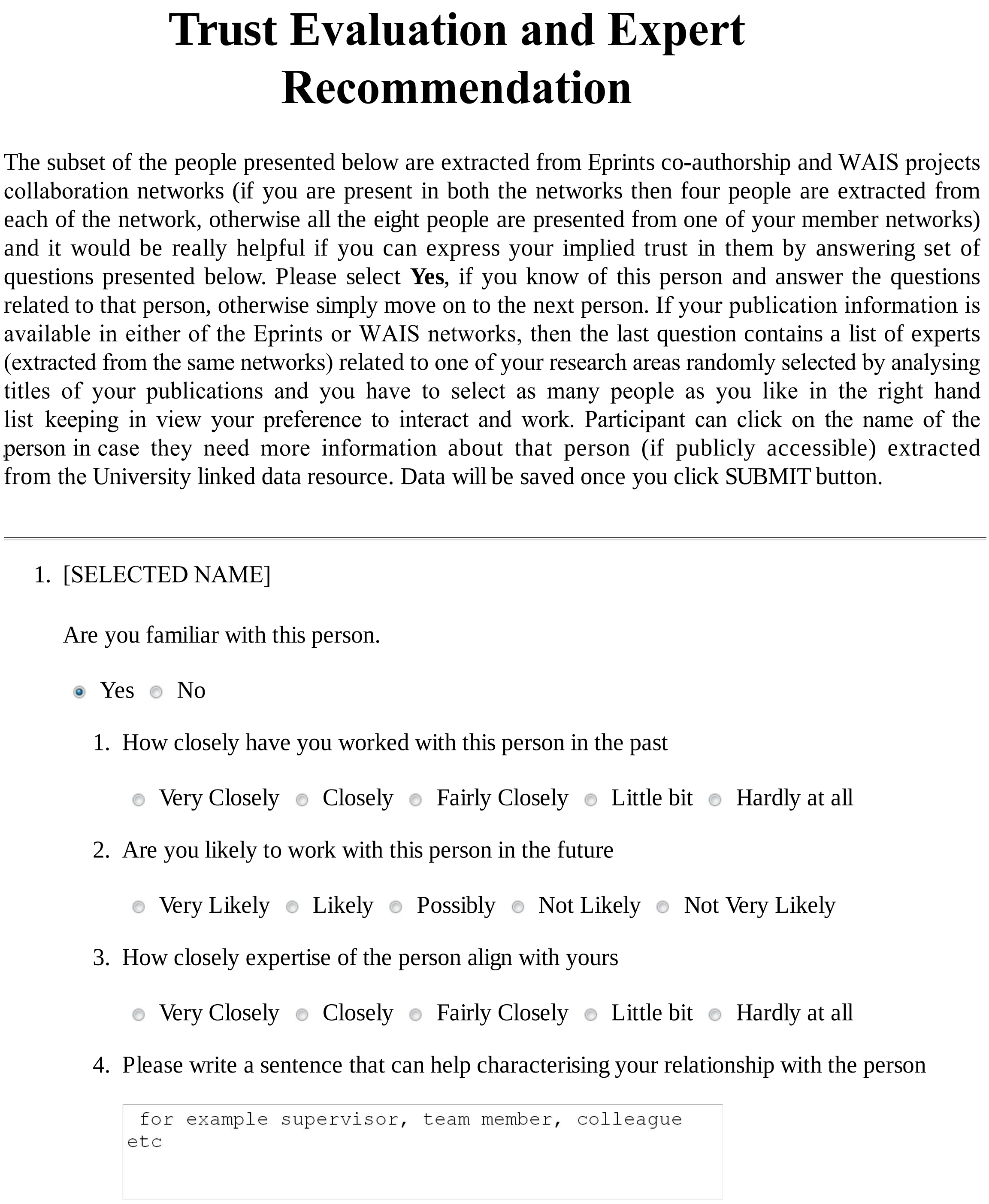
Login

Figure 6.3: Login page of a trust and expert recommendation survey appilcation

To access the actual survey page each *rating* participant had to log-in using *First* and *Last Name* attributes. Once successfully logged-in, a dynamic survey was generated for each participant with set of *eight* users to establish their ratings on professional trust, looking at the social graph assuming participant as an ego node. Due to the selection of *rated* participants from different path lengths from the *rating* participant it first asked the participant about familiarity and if participant knew that person, it asked a set of subsequent question about that person. A set of queries described in relevant sections

[2] https://www.djangoproject.com/

was sent to the Sesame SPARQL endpoint and that extracted the data that needs to be presented to that specific user. If the profile information of the *rated* participant was available on the University web server, it too was extracted using the linked data provided under the Open Data initiative here at University of Southamtpton.



Trust Evaluation and Expert Recommendation

The subset of the people presented below are extracted from Eprints co-authorship and WAIS projects collaboration networks (if you are present in both the networks then four people are extracted from each of the network, otherwise all the eight people are presented from one of your member networks) and it would be really helpful if you can express your implied trust in them by answering set of questions presented below. Please select **Yes**, if you know of this person and answer the questions related to that person, otherwise simply move on to the next person. If your publication information is available in either of the Eprints or WAIS networks, then the last question contains a list of experts (extracted from the same networks) related to one of your research areas randomly selected by analysing titles of your publications and you have to select as many people as you like in the right hand list keeping in view your preference to interact and work. Participant can click on the name of the person in case they need more information about that person (if publicly accessible) extracted from the University linked data resource. Data will be saved once you click SUBMIT button.

---

1. [SELECTED NAME]

   Are you familiar with this person.

   Yes No

   1. How closely have you worked with this person in the past

      Very Closely Closely Fairly Closely Little bit Hardly at all

   2. Are you likely to work with this person in the future

      Very Likely Likely Possibly Not Likely Not Very Likely

   3. How closely expertise of the person align with yours

      Very Closely Closely Fairly Closely Little bit Hardly at all

   4. Please write a sentence that can help characterising your relationship with the person

      for example supervisor, team member, colleague etc

Figure 6.4: Questions to extract professional trust between participants in survey application

6.3.2 Hosting and Execution

Before hosting or running any experiment that involves human participation and specifically in this case when the participation deems answering personal questions, an application bearing all the information about the experiment needs to be submitted to the Ethics Committee[3] for approval. This Committee scrutinised all the provided information, including snapshots of the survey, reviewed all five submitted documents - see Appendix C) and agreed that it followed the code of ethics according to its charter, awarding a four-digit ethics reference code to the survey, 4406, and gave approval to start. In doing so, it took full responsibility for the ethical issues and built confidence of the participants, giving them a forum to complain if any of the rules were breached by researcher or anything controversial happened in the meantime.

The survey was hosted at the virtual space on the University web server, URL (trustsurvey.ecs.soton.ac.uk/login/). The application was circulated among potential participants for a period of one month and their submitted data was stored in MySQL database.

6.4 Results and Analysis

Some 26 users participated in this survey experiment, each *rating* participants providing trust ratings about, on average, 3.15 participants out of an average of 5.38 *rated* participants presented. Table 6.18 shows the breakdown of the *trust ratings* corresponding to different *path lengths*. It shows that majority of the *rated* participants of *path length* 1 are answered for the *trust ratings*, but this number decreases as the *path length* increases. This is due to unfamiliarity between people because of absence of any direct interaction.

Table 6.2: Breakdown of trust ratings corresponding to path lengths of 1, 2 and 3

Path Length	Total Pairs - Don′t Know = Trust Ratings
1	55 - 6 = 49
2	45 - 27 = 18
3	40 - 25 = 15

.

6.4.1 Trust Data Description

As mentioned in Section 6.2.2, interpersonal trust data from networks and survey can be categorised as of three types due to users' presence in different regions of the consolidated

[3]https://www.ergo.soton.ac.uk/

networks (also shown as a Venn diagram in Figure 6.1). The first type are those with each member of the pair in both the networks, belonging to region EW, represented as $N_p^{overlapping}$ users; the second are cross-region users where each member of the pair belongs to different regions (E, W, EW) known as $N_p^{cross-region}$ users. The last type of users are those belonging to a single network, that is, both users in the pair belong to one of the regions (E or W), represented using $N_p^{single-network}$ users.

It is important to analyse trust from these different pairs of users because they represent different aggregation scenarios (discussed in Section 4.4.1). Comparing values corresponding to those obtained from the proxy trust survey helps to justify the third claim of the hypothesis.

Hypothesis: Trust metrics generated over MuDi social networks increase accuracy of trust over individual networks in terms of more accurately capturing how users perceive one another in real life.

Table 6.3: Data description for the trust parameters obtained from system and survey experiments. Team Member, ECS Colleague, WAIS Colleague and Supervisor are abbreviated as TM, EC, WC, SP

Data Nature	Trust Ratings	Range
System readings	Eprints Co-authorship proxy trust ($trust^{eprints}$)	(0,1)
	WAIS projects collaboration proxy trust ($trust^{wais}$)	(0,1)
	Consolidated proxy trust ($trust^{mudi}$)	(0,1)
Survey readings	Past proxy trust ($trust^{past}$)	[0.2, 0.4, 0.6, 0.8, 1.0]
	Future proxy trust ($trust^{future}$)	[0.2, 0.4, 0.6, 0.8, 1.0]
	Relationship (Rel)	[TM, EC, WC, SP]
	Expertise_Match (E_M)	[0.2, 0.4, 0.6, 0.8, 1.0]

There were two trust data points available from the survey experiment (represent *past trust* ($trust^{past}$) and *future trust* ($trust^{future}$) as shown in Table 6.3) for each category of user pairs described above, but there was a variable number of trust data points available from system readings. For $N_p^{overlapping}$ pairs, there are three system-generated trust data points from each of the trust propagation algorithms: two from each of the individual networks ePrints ($trust^{eprints}$) and WAIS ($trust^{wais}$), and one from the consolidated version of these MuDi ($trust^{mudi}$) networks.

For $N_p^{cross-region}$ and $N_p^{single-network}$ pairs, however, there are only two data points available, one from either of the individual networks ePrints ($trust^{eprints}$) or WAIS ($trust^{wais}$) and the other from the consolidated version of MuDi ($trust^{mudi}$) networks. $N_p^{cross-region}$ users for this experiment mostly contained those existing across EW $\rightarrow$ E and E $\rightarrow$ EW regions as articles published by most of those on WAIS projects also come under the ePrints network. This brought them into the region EW, while the other member of the pair was, for example, a PhD student having supervisee relation with the former and who has published a joint article together in the ePrints network.

$N_p^{single-network}$ users for this experiment were those belonging to E region and could be, for example, two PhD students who co-authred a paper. Here, participant pairs from EW $\rightarrow$ W, E $\rightarrow$ W (for $N_p^{cross-region}$ scenario) and W region (for $N_p^{single-network}$ scenario) are not considered for analysis, first because W region is approximately a subset of E and second because the rest of the those working on projects mostly come from industry or some another university and it was hard to contact them for survey. Also there was no consistent URI for them in the networks and system generated a new random URI each time they were mentioned in the RDF dataset, thus creating anomalies in the data.

Both system and survey readings belonging to different categories of users were scaled in the range (0,1). Data from the system was left skewed due to having high frequency of single co-authorship values between participants than rare high trust values. Scaling it by dividing maximum among all the co-authorship metrics could result in diminishing small trust values thus losing important trust information. To avoid the dampening, data was first normalised by taking log base 10 of all data points in both the individual networks. Then the maximum of all the values from that specific network acts as divisor to find scaled data in the range (0,1). Equation 6.2 explains the procedure.

$$\frac{log(T_{N_p}i)}{max(log(\forall \quad T_{N_p}j \quad \in \quad T_{N_p}))} \tag{6.2}$$

The survey data was a likert scale in the range 1 to 5, the scaling of the data was not going to dampen small values. Hence the Equation 6.3 was used to scale in the range (0,1).

$$\frac{T_{N_p}i}{5} \tag{6.3}$$

Other than this data, there were two additional statements about the type of relationship (Rel) and expertise match (E_M) between each pair of participants that would help us analyse accuracy of aggregated trust metric with respect to Rel and E_M. Survey readings are discrete values in the range [0, 0.8] with increments of 0.2 except Rel data, which contains four categories: Team Member (TM); ECS Colleague (EC); WAIS colleague (WC); and Supervisor (SP). System trust readings are continuous values in the range (0,1).

6.4.1.1 Overlapping users' data

Trust data from overlapping users ($N_p^{overlapping}$) (region EW in Figure 6.1) is shown in Table B.2 (see Appendix B) for two different trust algorithms and two proxy trust questions.

Looking at the data with respect to path length (PL) of 1 shows that past trust (represented using proxy value *past work* ($trust^{past}$)) and future trust, (represented using proxy value *future work* ($trust^{future}$)) for most of the data reside either close to ePrints or WAIS with MuDi always giving the version of trust closer to $trust^{past}$ or $trust^{future}$ in each case, compared to one of the individual networks. In 39% cases MuDi gave closer trust values than both other networks (i.e. data points 5, 7, 8, 23 for $trust^{past}$ and 2, 5, 7, 12, 16, 19, 20, 23 for $trust^{past}$), but this was a minority case and the reason for such results is described in analysis (Section 6.4.2). For this scenario, both trust algorithms gave same value of trust for $N_p^{overlapping}$ users as they return directly aggregated trust values between them.

For PL of 2, MuDi values were also closer to $trust^{past}$ and $trust^{future}$ than any one of the individual network in each case, and here trust propagation using shortest path appeared to behave better than the strongest path. In 25% cases (ratings of 26 and 29 for $trust^{future}$) MuDi gives trust estimations that were better than either networks.

Simialrly, for PL of 3, MuDi approached closer to $trust^{past}$ and $trust^{future}$ than one of the individual networks in all cases, and in 33% cases (data points 36 and 37 for $trust^{past}$ and $trust^{future}$), it shows better values than both of the individual networks. In terms of searching trust paths, apparently the shortest trust path algorithm gives results that are closer to actual values $trust^{past}$ and $trust^{future}$, and the reason of being this may be the sharp decay of professional trust between people in real life. Reviewing the data for *Rel* shows that MuDi system data points for TM and WC seem to approach real life values closer than SP relation, and the reason is discussed in Section 6.4.2

6.4.1.2 Cross-region users' Data

Trust data for cross-network users pairs ($N_p^{cross-region}$) (EW $\rightarrow$ E and E $\rightarrow$ EW links) using the same parameters as of the overlapping users is shown in Table B.3 (see Appendix B). As one member of the pair also belonged to another region, so there was only single trust value available between them.

Trust values for path length (PL) of 1 in MuDi networks remained approximately the same (with minute difference) as individual networks due to the availability of only a single value from one of the networks, preserving the integrity of trust mentioned during the simulation experiment. However, that difference showed a slight deterioration of trust in consolidated networks compared to individual networks, and in only 26% cases did trust from MuDi networks end up closer to real trust values (for data points 4, 15, 16, 20, 21, 23 w.r.t $trust^{future}$). A detailed anaylsis of this behaviour is undertaken in Section 6.4.2, but apparently this was due to the nature of different relations (i.e. *Rel*) between users in these networks.

For PL of 2, MuDi values were near to either $trust^{past}$ or $trust^{future}$ in all cases as compared to those from individual networks, while for PL of 3, this happened in 25% of the cases. In both these cases the shortest path algorithm gave either similar or results similar to that of strongest path algorithm.

6.4.1.3 Single-network users' data

Trust data between users belonging to individual social networks $N_p^{single-network}$ (those in regions E), from individual as well as MuDi networks, is shown in Table B.4 (see Appendix B). As these users are part of single networks there is only one trust metric available between them and the nature of this data is similar to that of $N_p^{cross-region}$ users.

Looking at the data for path length (PL) of 1 shows that MuDi values were slightly different from those of individual networks as, like $N_p^{cross-region}$ users, it also considers unavailability of trust value from one of the network due to absence of trust. In 33% of the cases, this difference brought MuDi values closer to $trust^{future}$, and in other 67% cases, it carried the trust away from $trust^{past}$ and $trust^{future}$ by a small difference. For PL of 2, MuDi brought trust values closer to real values in 33% of the cases, while for PL of 3, this was the case 80% of the time. Results also show that data for PL of 2 and 3 from shortest path algorithm gave better results than the strongest path algorithm.

6.4.2 Trust Data Analysis

The results of this experiment can be analysed in two scenarios; first we analysed statistically whether the aggregated trust data from MuDi networks ($trust^{mudi}$) were more *similar* to real trust metrics ($trust^{past}$, $trust^{future}$) than each of the individual networks ($trust^{eprints}$, $trust^{wais}$). Second we studied statistically whether the generated trust metrics from MuDi networks ($trust^{mudi}$) brought trust metrics *closer* to real trust metrics ($trust^{past}$, $trust^{future}$) compared to those from individual networks ($trust^{eprints}$, $trust^{wais}$). Both the tests are conducted by evaluating p-value between datasets using T-Test, but for the former scenario, it was analysed between each of the corresponding systems and survey readings, while in the latter case, it was conducted between the absolute difference of system and survey readings.

6.4.2.1 Statistical Significance test to analyse the similarity between system and survey readings

The statistical test for the first scenario is a two-tailed paired T-Test, aiming to find whether the trust metrics generated for same pair of participants by MuDi networks are more similar to those given by $trust^{past}$ and $trust^{future}$ than those from individual networks ePrints and WAIS. It generates a p-value and if the following null hypothesis is true for MuDi trust metrics, our claim that aggregated MuDi trust metrics are the one more similar to $trust^{past}$ and $trust^{future}$ is proven.

Hypothesis: If the value of $p \leq 0.05$, this means system generated trust metrics are significantly *dissimilar* from survey generated trust metrics.

Null Hypothesis: If the value of $p > 0.05$, this means system generated trust metrics are not significantly *dissimilar* from survey generated trust metrics.

This statistical significance was tested for the three categories of dataset mentioned in Sections 6.4.1.1, 6.4.1.2 and 6.4.1.3.

First, by analysing the p-value for overlapping users ($N_p^{overlapping}$) presented in Table 6.4 and Table 6.5, it can be seen that its value for ePrints ($trust^{eprints}$) and WAIS ($trust^{wais}$) for both the algorithms is statistically significantly in comparison with both $trust^{past}$ and $trust^{future}$ i.e. $p \leq 0.05$ in all the cases, which proves that these two pairs of values are significantly dissimilar. But when analysing same p-value for MuDi networks ($trust^{mudi}$) and $trust^{past}/trust^{future}$, it is no longer statistically significant for dissimilarity. Its value becomes $p > 0.05$, which actually proves null hypothesis, and shows that the $trust^{mudi}$ is similar to those from real life (both $trust^{past}$ and $trust^{future}$).

Table 6.4: Statistical significance (p-value) evaluation of trust data *similarity* between system and survey readings for overlapping users (strongest path algo)

System Readings / *Survey Readings*	Strongest Path Algo		
	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$
$trust^{past}$	0.01	0.01	0.13
$trust^{future}$	0.01	0.01	0.14

Table 6.5: Statistical significance (p-value) evaluation of trust data *similarity* between system and survey readings for overlapping users (shortest path algo)

System Readings / *Survey Readings*	Shortest Path Algo		
	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$
$trust^{past}$	< 0.01	0.01	0.06
$trust^{future}$	< 0.01	< 0.01	0.06

The p-value for cross-region users ($N_p^{cross-region}$) is shown in Table 6.6 and the results show that MuDi networks did not bring any difference to the dissimilarity of trust values between system and survey readings. Individual trust metrics from ePrints (i.e. ($trust^{eprints}$)) which were significantly dissimilar from ($trust^{past}$) and ($trust^{future}$), remain significantly dissimilar when evaluated over MuDi social networks. Hence the null hypothesis is not true in this case and the hypothesis stands true that system-generated trust metrics for both individual and MuDi networks are dissimilar to survey readings.

Table 6.6: Statistical significance (p-value) evaluation of trust data *similarity* between system and survey readings for cross-region users

System Readings / *Survey Readings*	Strongest Path Algo		Shortest Path Algo	
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$
$trust^{past}$	< 0.01	< 0.01	< 0.01	< 0.01
$trust^{future}$	< 0.01	< 0.01	< 0.01	< 0.01

Unlike both $N_p^{overlapping}$ and $N_p^{cross-region}$ users, the p-value for single- network ($N_p^{single-network}$) users behaves differently for $trust^{past}$ and $trust^{future}$ as shown in Table 6.7. For $trust^{past}$, $p \leq 0.05$, which proves the hypothesis that the statistical significance for dissimilarity holds true, this means that none of the individual and MuDi trust metrics are similar to real trust metric $trust^{past}$. For $trust^{future}$ however, $p > 0.05$ shows a statistically significant similarity between system and survey readings for both individual and MuDi networks. Hence, the hypothesis stands true for $trust^{past}$, while the null hypothesis is true for $trust^{future}$.

Table 6.7: Statistical significance (p-value) evaluation of trust data *similarity* between system and survey readings for single-network users

System Readings / *Survey Readings*	Strongest Path Algo		Shortest Path Algo	
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$
$trust^{past}$	0.02	0.02	0.02	0.02
$trust^{future}$	0.40	0.33	0.38	0.31

From this analysis, it is proven that trust metrics from consolidated MuDi networks are significantly similar to those obtained from real life only for only $N_p^{overlapping}$ users; otherwise it remains dissimilar in both the other categories of users apart from $N_p^{single-network}$ users when system readings are analysed with the $trust^{future}$ survey metric.

To analyse whether comparative similarity between trust metrics from MuDi networks and $trust^{past}$/ $trust^{future}$ actually results in bringing trust metrics closer, the next set of analyses calculates the difference between system and survey readings and then again runs a statistical significance test to find whether the difference between MuDi readings and $trust^{past}$/ $trust^{future}$ is significantly less than the difference between $trust^{eprints}$/ $trust^{wais}$ and $trust^{past}$/ $trust^{future}$.

6.4.2.2 Absolute difference between System and Survey Readings

The analysis conducted in Section 6.4.2.1 has helped us understanding whether the consolidation of MuDi networks results in trust metrics that are similar to real life trust, as we claimed. We proved to be true for overlapping users, while false for cross-region and those present in only single social network.

To analyse whether this similarity of MuDi trust metrics with respect to $trust^{past}$ and $trust^{future}$ actually brings MuDi values significantly *closer* to $trust^{past}$ and $trust^{future}$ compared to those from ePrints and WAIS, this set of analyses runs statistical significance test over absolute difference between system and survey readings. Table 6.8, 6.9, 6.10 present absolute differences between those readings.

Table 6.8: Absolute difference between system and survey trust readings for overlapping users. EP represents difference between trust readings $trust^{eprints}$ and $trust^{past}$, WP shows difference between$trust^{wais}$ and $trust^{past}$ and MP shows difference between $trust^{mudi}$ and $trust^{past}$. Similarly EF shows difference column between $trust^{eprints}$ and $trust^{future}$, WF shows difference column between $trust^{wais}$ and $trust^{future}$ and MF between $trust^{mudi}$ and $trust^{future}$.

No	Strongest Path Readings						Shortest Path Readings					
	EP	WP	MP	EF	WF	MF	EP	WP	MP	EF	WF	MF
1	0.65	0.73	0.66	0.25	0.33	0.26	0.65	0.73	0.66	0.25	0.33	0.26
2	0.50	0.74	0.52	0.10	0.14	0.08	0.50	0.74	0.52	0.10	0.14	0.08
3	0.33	0.53	0.35	0.13	0.33	0.15	0.33	0.53	0.35	0.13	0.33	0.15
4	0.28	0.37	0.29	0.28	0.37	0.29	0.28	0.37	0.29	0.28	0.37	0.29
5	0.05	0.01	0	0.05	0.01	0	0.05	0.01	0	0.05	0.01	0
6	0.42	0.34	0.41	0.62	0.54	0.61	0.42	0.34	0.41	0.62	0.54	0.61
7	0.26	0.13	0.09	0.26	0.13	0.09	0.26	0.13	0.09	0.26	0.13	0.09
8	0.12	0.03	0.02	0.32	0.17	0.18	0.12	0.03	0.02	0.32	0.17	0.18
9	0.05	0.07	0.06	0.25	0.13	0.14	0.05	0.07	0.06	0.25	0.13	0.14
10	0.10	0.13	0.10	0.30	0.33	0.30	0.10	0.13	0.10	0.30	0.33	0.30
11	0.45	0.33	0.34	0.65	0.53	0.54	0.45	0.33	0.34	0.65	0.53	0.54
12	0.1	0.07	0.10	0.3	0.10	0.13	0.1	0.07	0.10	0.3	0.10	0.13
13	0.25	0.17	0.18	0.25	0.17	0.18	0.25	0.17	0.18	0.25	0.17	0.18
14	0.62	0.53	0.54	0.22	0.13	0.14	0.62	0.53	0.54	0.22	0.13	0.14
15	0.26	0.13	0.14	0.46	0.33	0.34	0.26	0.13	0.14	0.46	0.33	0.34
16	0.18	0.13	0.13	0.02	0.07	0.07	0.18	0.13	0.13	0.02	0.07	0.07
17	0.47	0.33	0.34	0.27	0.13	0.14	0.47	0.33	0.34	0.27	0.13	0.14
18	0.10	0.13	0.10	0.30	0.33	0.30	0.10	0.13	0.10	0.30	0.33	0.30

EP $= |trust^{eprints} - trust^{past}|$ EF $= |trust^{eprints} - trust^{future}|$
WP $= |trust^{wais} - trust^{past}|$ WF $= |trust^{wais} - trust^{future}|$
MP $= |trust^{mudi} - trust^{past}|$ MF $= |trust^{mudi} - trust^{future}|$

Continued on next page

Table 6.8 – continued from previous page

No	Strongest Path Readings						Shortest Path Readings					
	EP	WP	MP	EF	WF	MF	EP	WP	MP	EF	WF	MF
19	0.05	0.23	0.20	0.25	0.03	0	0.05	0.23	0.20	0.25	0.03	0
20	0.45	0.06	0.10	0.25	0.14	0.10	0.45	0.06	0.10	0.25	0.14	0.10
21	0.45	0.53	0.46	0.45	0.53	0.46	0.45	0.53	0.46	0.45	0.53	0.46
22	0.36	0.33	0.33	0.36	0.33	0.33	0.36	0.33	0.33	0.36	0.33	0.33
23	0.10	0.13	0.08	0.10	0.13	0.08	0.10	0.13	0.08	0.10	0.13	0.08
24	0.23	0.65	0.28	0.03	0.45	0.08	0.23	0.65	0.28	0.03	0.45	0.08
25	0.05	0.08	0.06	0.45	0.48	0.46	0.05	0.13	0.06	0.45	0.53	0.46
26	0.12	0.13	0.13	0.12	0.13	0.13	0.11	0.12	0.09	0.11	0.12	0.09
27	0.31	0.17	0.33	0.31	0.17	0.33	0.31	0.17	0.33	0.31	0.17	0.33
28	0.15	0.19	0.14	0.05	0.01	0.06	0.15	0.12	0.14	0.05	0.08	0.06
29	0.17	0.17	0.12	0.17	0.17	0.12	0.21	0.12	0.08	0.21	0.12	0.08
30	0.07	0.12	0.12	0.07	0.12	0.12	0.04	0.12	0.07	0.04	0.12	0.07
31	0.18	0.17	0.17	0.18	0.17	0.17	0.18	0.17	0.17	0.18	0.17	0.17
32	0.10	0.04	0.08	0.1	0.04	0.08	0.10	0.04	0.08	0.10	0.04	0.08
33	0.17	0.10	0.20	0.17	0.10	0.20	0.06	0.10	0.07	0.06	0.1	0.07
34	0.06	0.19	0.08	0.26	0.39	0.28	0.09	0.19	0.12	0.29	0.39	0.32
35	0.37	0.11	0.37	0.37	0.11	0.37	0.21	0.11	0.22	0.21	0.11	0.22
36	0.12	0.17	0.13	0.12	0.17	0.13	0.126	0.18	0.13	0.16	0.18	0.13
37	0.05	0.17	0.05	0.05	0.17	0.05	0.05	0.18	0.02	0.05	0.18	0.02

EP = $|trust^{eprints} - trust^{past}|$ EF = $|trust^{eprints} - trust^{future}|$
WP = $|trust^{wais} - trust^{past}|$ WF = $|trust^{wais} - trust^{future}|$
MP = $|trust^{mudi} - trust^{past}|$ MF = $|trust^{mudi} - trust^{future}|$

Table 6.9: Absolute difference between system and survey trust readings for cross-region users. EP represents difference between trust readings $trust^{eprints}$ and $trust^{past}$, and MP shows difference between $trust^{mudi}$ and $trust^{past}$. Similarly EF shows difference column between $trust^{eprints}$ and $trust^{future}$, and MF between $trust^{mudi}$ and $trust^{future}$.

No	Strongest Path Readings				Shortest Path Readings			
	EP	MP	EF	MF	EP	MP	EF	MF
1	0.56	0.58	0.16	0.18	0.56	0.58	0.16	0.18
2	0.56	0.58	0.56	0.58	0.56	0.58	0.56	0.58
3	0.65	0.66	0.25	0.26	0.65	0.66	0.25	0.26
4	0.56	0.58	0.24	0.22	0.56	0.58	0.24	0.22
5	0.37	0.41	0.17	0.21	0.37	0.41	0.17	0.21

EP = $|trust^{eprints} - trust^{past}|$ EF = $|trust^{eprints} - trust^{future}|$
MP = $|trust^{mudi} - trust^{past}|$ MF = $|trust^{mudi} - trust^{future}|$

Continued on next page

Table 6.9 – continued from previous page

No	Strongest Path Readings				Shortest Path Readings			
	EP	MP	EF	MF	EP	MP	EF	MF
6	0.56	0.58	0.56	0.58	0.56	0.58	0.56	0.58
7	0.56	0.58	0.56	0.58	0.56	0.58	0.56	0.58
8	0.45	0.48	0.05	0.08	0.45	0.48	0.05	0.08
9	0.56	0.58	0.36	0.38	0.56	0.58	0.36	0.38
10	0.25	0.26	0.25	0.26	0.25	0.26	0.25	0.26
11	0.41	0.45	0.41	0.45	0.41	0.45	0.41	0.45
12	0.5	0.53	0.5	0.53	0.5	0.53	0.5	0.53
13	0.56	0.58	0.24	0.22	0.56	0.58	0.24	0.22
14	0.45	0.46	0.25	0.26	0.45	0.46	0.25	0.26
15	0.65	0.66	0.15	0.14	0.65	0.66	0.15	0.14
16	0.1	0.07	0.1	0.07	0.1	0.07	0.1	0.07
17	0.42	0.44	0.42	0.44	0.42	0.44	0.42	0.44
18	0.38	0.42	0.18	0.22	0.38	0.42	0.18	0.22
19	0.45	0.46	0.25	0.26	0.45	0.46	0.25	0.26
20	0.45	0.46	0.05	0.06	0.45	0.46	0.05	0.06
21	0.17	0.19	0.23	0.21	0.17	0.19	0.23	0.21
22	0.25	0.26	0.25	0.26	0.25	0.26	0.25	0.26
23	0.37	0.41	0.43	0.39	0.37	0.41	0.43	0.39
24	0.1	0.1	0.1	0.1	0.1	0.09	0.1	0.09
25	0.11	0.1	0.09	0.1	0.09	0.08	0.11	0.12
26	0.04	0.05	0.16	0.15	0.1	0.11	0.1	0.09
27	0.09	0.07	0.11	0.13	0.1	0.07	0.1	0.13
28	0.1	0.13	0.1	0.13	0.08	0.07	0.08	0.07
29	0.07	0.06	0.07	0.06	0.07	0.06	0.07	0.06
30	0.08	0.07	0.08	0.07	0.06	0.05	0.06	0.05
31	0.32	0.33	0.32	0.33	0.32	0.37	0.32	0.37
32	0.02	0.09	0.18	0.11	0.02	0.05	0.18	0.15
33	0.02	0.02	0.18	0.22	0.08	0.17	0.28	0.37
34	0.05	0.05	0.15	0.15	0.04	0.04	0.16	0.16

EP = $|trust^{eprints} - trust^{past}|$ EF = $|trust^{eprints} - trust^{future}|$

MP = $|trust^{mudi} - trust^{past}|$ MF = $|trust^{mudi} - trust^{future}|$

Table 6.10: Absolute difference between system and survey trust readings for single-network users. EP represents difference between trust readings $trust^{eprints}$ and $trust^{past}$ and MP shows difference between $trust^{mudi}$ and $trust^{past}$. Similarly EF shows difference column between $trust^{eprints}$ and $trust^{future}$, and MF between $trust^{mudi}$ and $trust^{future}$.

No	Strongest Path Readings				Shortest Path Readings			
	EP	MP	EF	MF	EP	MP	EF	MF
1	0.50	0.53	0.10	0.13	0.50	0.53	0.10	0.13
2	0.36	0.38	0.24	0.22	0.36	0.38	0.24	0.22
3	0.45	0.46	0.45	0.46	0.45	0.46	0.45	0.46
4	0.57	0.57	0.03	0.03	0.58	0.58	0.02	0.02
5	0.38	0.38	0.58	0.58	0.33	0.34	0.53	0.54
6	0.06	0.04	0.06	0.04	0.04	0.03	0.04	0.03
7	0.16	0.18	0.04	0.02	0.17	0.19	0.03	0.01
8	0.01	0.01	0.01	0.01	0.01	0	0.01	0
9	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
10	0.02	0.01	0.02	0.01	0.01	0	0.01	0
11	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01
EP = $\lvert trust^{eprints} - trust^{past}\rvert$ EF = $\lvert trust^{eprints} - trust^{future}\rvert$								
MP = $\lvert trust^{mudi} - trust^{past}\rvert$ MF = $\lvert trust^{mudi} - trust^{future}\rvert$								

6.4.2.3 Statistical significance test to analyse the closeness between system and survey readings

The statistical significance test for this analysis is a one-tailed paired T-Test to prove whether trust metrics from MuDi networks actually improves the trust calculations by bringing trust values *closer* to $trust^{past}$ and $trust^{future}$ estimating the p-value for it. If the p-value follows the hypothesis mentioned below, then it shows that MuDi actually improves trust calculations, otherwise it does not improves significantly on the results from individual networks.

Hypothesis: If p $\leq$ 0.05, this shows significant improvement of MuDi trust values over individual networks in terms of *closeness* to $trust^{past}$ and $trust^{future}$.

Null Hypothesis: If p $>$ 0.05, there is no significant improvement of MuDi trust values over individual networks in terms of *closeness* to $trust^{past}$ and $trust^{future}$.

There is an issue with the use of the one-tailed T-Test here; it only helps us in estimating the significant absolute difference between the two datasets using the p-value. This difference between system and survey readings, in this case, can be due to two reasons.

Either it can due to improvement of trust values from consolidated networks, that is, MP < EP (for example, see data point *11* in Table 6.8), or it can be due to deterioration of trust values from consolidated MuDi networks, that is, MP > EP (for example, see data point *11* in Table 6.9). Here MP = $|trust^{mudi} - trust^{past}|$ and EP = $|trust^{eprints} - trust^{past}|$. To analyse whether this significant difference is due to improvement, p-value needs to be coupled with the *mean* value of the corresponding data (presented in Tables 6.11 and 6.12).

If p $\leq$ 0.05 and {*mean*(MP) < *mean*(EP) & *mean*(MP) < mean(WP)} or p $\leq$ 0.05 and {*mean*(MF) < *mean*(EF) & *mean*(MF) < mean(WF)}, this means the significant difference resulted in improvement of trust metrics and if p $\leq$ 0.05 and {*mean*(MP) > *mean*(EP) & *mean*(MP) > mean(WP)} or p $\leq$ 0.05 and {*mean*(MF) > *mean*(EF) & *mean*(MF) > mean(WF)}, this means the statistically significant absolute difference between datasets has resulted in deterioration of trust metrics. So, in both the cases, although there is a statistically significant difference i.e. p $\leq$ 0.05, because in the latter case that difference is due to the increase in *mean* of difference between $trust^{mudi}$ and $trust^{past}/trust^{future}$ compared to the *mean* of difference between $trust^{eprints}/trust^{wais}$ and $trust^{past}/trust^{future}$, it represents deterioration of trust metrics after consolidating MuDi social networks. The hypothesis specified above can be further refined and re-written as below:

Hypothesis: If p $\leq$ 0.05 and {*mean*(MP) < *mean*(EP) and *mean*(MP) < *mean*(WP)} or p $\leq$ 0.05 and {*mean*(MF) < *mean*(EF) and *mean*(MF) < *mean*(WF)}, this shows significant improvement of MuDi trust values over individual networks in terms of *closeness* to $trust^{past}$ and $trust^{future}$.

Null Hypothesis: If p $\leq$ 0.05 and {*mean*(MP) > *mean*(EP) and *mean*(MP) > *mean*(WP)} or p $\leq$ 0.05 and {*mean*(MF) > *mean*(EF) and *mean*(MF) > *mean*(WF), there is no significant improvement of MuDi trust values over individual networks in terms of *closeness* to $trust^{past}$ and $trust^{future}$.

Keeping this analogy in mind, and first analysing it for overlapping ($N_p^{overlapping}$) users, Tables 6.11 and 6.12 show that for both strongest and shortest path algorithms *mean*(MP) < *mean*(EP) and *mean*(MP) < mean(WP). Similarly *mean*(MF) < *mean* (EF) and *mean*(MF) < mean(WF) (both highlighted in **bold**). This means that for $N_p^{overlapping}$ users, absolute difference between trust readings from MuDi networks and survey is less than the difference between ePrints or WAIS and survey. However, for cross-region and single-network users, *mean*(MP) $\geq$ *mean*(EP) and *mean*(MF) $\geq$ mean(WF) (highlighted in **italic**), which means that the difference between trust values from MuDi networks and survey has increased. As a result trust metrics from individual network ePrints were closer to survey readings than MuDi networks.

Table 6.11: Mean (M) of the difference of system and survey readings for strongest path algorithm present in Tables 6.8, 6.9, 6.10

	Strongest Path Algo					
	EP	WP	MP	EF	WF	MF
$N_p^{overlapping}$	**0.24**	**0.23**	**0.21**	**0.23**	**0.22**	**0.20**
	EP		MP	EF		MF
$N_p^{cross-region}$	*0.33*		*0.35*	*0.24*		*0.25*
$N_p^{single-network}$	*0.23*		*0.24*	*0.14*		*0.14*
EP $= \vert trust^{eprints} - trust^{past}\vert$			EF $= \vert trust^{eprints} - trust^{future}\vert$			
WP $= \vert trust^{wais} - trust^{past}\vert$			WF $= \vert trust^{wais} - trust^{future}\vert$			
MP $= \vert trust^{mudi} - trust^{past}\vert$			MF $= \vert trust^{mudi} - trust^{future}\vert$			

Table 6.12: Mean (M) of the difference of system and survey readings for shortest path algorithm present in Tables 6.8, 6.9, 6.10

	Shortest Path Algo					
	EP	WP	MP	EF	WF	MF
$N_p^{overlapping}$	**0.23**	**0.23**	**0.20**	**0.23**	**0.22**	**0.19**
	EP		MP	EF		MF
$N_p^{cross-region}$	*0.33*		*0.35*	*0.24*		*0.25*
$N_p^{single-network}$	*0.23*		*0.23*	*0.13*		*0.13*
EP $= \vert trust^{ePrints} - trust^{past}\vert$			EF $= \vert trust^{ePrints} - trust^{future}\vert$			
WP $= \vert trust^{wais} - trust^{past}\vert$			WF $= \vert trust^{wais} - trust^{future}\vert$			
MP $= \vert trust^{mudi} - trust^{past}\vert$			MF $= \vert trust^{mudi} - trust^{future}\vert$			

To further analyse whether the apparent decrease in mean difference between MuDi and survey readings for $N_p^{overlapping}$ users is statistically significant, Table 6.13 shows p-value between the absolute difference of system and survey readings presented in Table 6.8. Results show that for shortest path algorithm, p $\leq$ 0.05 for all the cases (highlighted in **bold**), while for strongest path algorithm p $>$ 0.05 between MP and WP, and then between MF and WF - see Table 6.13 for description of MP, WP, MF and WF. This shows that our hypothesis of MuDi metrics being closer to survey metrics for $N_p^{overlapping}$ users stands true for shortest path algorithm.

When analysing p-value for cross-region ($N_p^{cross-region}$) users (Table 6.14), for both strongest and shortest algorithms, p $\leq$ 0.05 (shown as **italics**), which shows that difference between trust metrics from MuDi networks and survey is statistically significant than those from ePrints. But if coupled with the mean of differences discussed above, that is, mean(MP) $>$ mean (EP) and mean(MP) $>$ mean(EP) (shown as **italic** in Tables 6.11 and 6.12), it proves that statistical significance does not show improvement, but a deterioration due to punishment of trust metrics from ePrints. Hence, null hypothesis stands true for this case.

Table 6.13: Statistical significance of *closeness* between system and survey readings for overlapping ($N_p^{overlapping}$) users.

.

$N_p^{overlapping}$ *Results* / *MuDi Results*	Strongest Path Algo		Shortest Path Algo	
	EP	WP	EP	WP
MP	0.03	0.09	**0.01**	**0.02**
	EF	WF	EF	WF
MF	< 0.01	0.13	$<$ **0.01**	**0.01**

EP $= \|trust^{eprints} - trust^{past}\|$ EF $= \|trust^{eprints} - trust^{future}\|$
WP $= \|trust^{wais} - trust^{past}\|$ WF $= \|trust^{wais} - trust^{future}\|$
MP $= \|trust^{mudi} - trust^{past}\|$ MF $= \|trust^{mudi} - trust^{future}\|$

Table 6.14: Statistical significance (p-value) of *closeness* between system and survey readings for *cross-region* ($N_p^{cross-region}$) users.

.

$N_p^{cross-region}$ *Results* / *MuDi Results*	Strongest Path Algo		Shortest Path Algo	
	EP	EF	EP	EF
MP	*< 0.01*	-	*< 0.01*	-
MF	-	*0.05*	-	*0.02*

EP $= \|trust^{eprints} - trust^{past}\|$ EF $= \|trust^{eprints} - trust^{future}\|$
MP $= \|trust^{mudi} - trust^{past}\|$ MF $= \|trust^{mudi} - trust^{future}\|$

When analysing single network ($N_p^{single-network}$) users, p-value > 0.05 for both strongest and shortest path algorithms (highlighted in **italic** in Table 6.15), which means null hypothesis stands true and consolidation of MuDi networks does not result in better values of trust for $N_p^{single-network}$ users. However, when coupled with mean difference from Tables 6.11 and 6.12, mean (EP) = mean (MP) and mean (EF) = mean (MF) (highlighted in **italic** in Tables 6.11 and 6.12) for both the strongest and shortest path algorithms, which shows that unlike cross-region it does not deteriorates trust metrics either, rather keeping it the same as in individual networks.

Table 6.15: Statistical significance (p-value) of *closeness* between syatem and survey readings for single network ($N_p^{single-network}$) users.

.

$N_p^{single-network}$ *Results* / *MuDi Results*	Strongest Path Algo		Shortest Path Algo	
	EP	EF	EP	EF
MP	*0.15*		*0.07*	-
MF	-	*0.27*	-	*0.38*

EP $= \|trust^{eprints} - trust^{past}\|$ EF $= \|trust^{eprints} - trust^{future}\|$
MP $= \|trust^{mudi} - trust^{past}\|$ MF $= \|trust^{mudi} - trust^{future}\|$

6.4.2.4 Discussion

Tables 6.16 and 6.17 summarise the results of claim 1.3 of the hypothesis. They show that for overlapping ($N_p^{overlapping}$) users, trust metrics generated from consolidated MuDi networks show improved behaviour for both $trust^{past}$ and $trust^{future}$ real metrics while for *cross-region* ($N_p^{cross-region}$) users, it deteriorates the trust metrics and hence trust values from individual networks are better than consolidated networks. For single network ($N_p^{single-network}$) users, it also does not improve the trust metrics but, unlike $N_p^{cross-region}$ users, does not deteriorates them; rather, trust metrics remain the same as from individual networks, so consolidation is neither beneficial nor detrimental for $N_p^{single-network}$ users.

Table 6.16: Outcome of the real world trust analysis when evaluated for $N_p^{overlapping}$, $N_p^{cross-region}$ and $N_p^{single-network}$ users in individual and consolidated MuDi networks in comparison with the *past work* ($trust^{past}$) survey trust question. ePrints represents trust values available from ePrints network, WAIS from WAIS projects collaboration network and MuDi from consolidated version of ePrints and WAIS networks.

Result / *Pair types*	Section 6.4.2.1			Section 6.4.2.3
	ePrints *similar*	WAIS *similar*	MuDi *similar*	MuDi *closer*
$N_p^{overlapping}$	×	×	✓	✓
	ePrints *similar*		MuDi *similar*	MuDi *closer*
$N_p^{cross-region}$	×		×	×
$N_p^{single-network}$	×		×	×

Table 6.17: Outcome of the real world trust analysis when evaluated for $N_p^{overlapping}$, $N_p^{cross-region}$ and $N_p^{single-network}$ users in individual and consolidated MuDi networks in comparison with the *future work* ($trust^{future}$) survey trust question. ePrints represents trust metrics available from ePrints network, WAIS from WAIS projects network and MuDi from consolidated version of ePrints and WAIS networks.

Result / *Pair types*	Section 6.4.2.1			Section 6.4.2.3
	ePrints *similar*	WAIS *similar*	MuDi *similar*	MuDi *closer*
$N_p^{overlapping}$	×	×	✓	✓
	ePrints *similar*		MuDi *similar*	$trust^{mudi}$ *closer*
$N_p^{cross-region}$	×		×	×
$N_p^{single-network}$	✓		✓	×

To analyse the reasons behind the results of this experiment better, the performance of the two propagation algorithms and data from *Rel* and *PL* parameters should also be reviewed. *Rel* shows the nature of relationship between each of the participant pair (N_p) that is analysed for trust metrics, and *PL* presents the length of the trust path they are connected with.

Looking into the results corresponding to each of the propagation algorithms, for $N_p^{overlapping}$ users only shortest path algorithm shows improvement, which means that the propagation decay of the trust along paths is very sharp. People trust indirectly connected people recommended by directly trusted friends, but the level of that trust decays rapidly with an increase in length of path, so the choice of finding strongest path for recommendation is not the right one.

Second, by looking at Tables B.2, B.3 and B.4 it can be seen that the most deteriorated consolidated metric with respect to *Rel* is the supervisor (SP) relationship, without even a single improved trust metric for all categories of users (i.e. $N_p^{overlapping}$, $N_p^{cross-region}$ and $N_p^{single-network}$). The reason for such behaviour is the high trust of students in their supervisors, while publishing far fewer articles (the metric by which to evaluate trust, in our system) due to being beginners in their field. Students rate their supervisors as highly trusted due to regular set of meetings and a series of continuous discussions throughout their PhD, for example, even if they have not published many articles during that time. Accordingly, when asked about their views relating to proxy trust questions, they rated their supervisors as highly trusted people academically in spite of not being involved to the same extent in terms of the numbers of publications. The most rational results were obtained from pairs of participants with a Team Member (TM) or WAIS colleague (WC) relationship, because the frequency of their collaboration, obtained from the ePrints and WAIS collaboration networks, is inline with that gathered from the real life professional trust by means of the survey.

With respect to PL, results are analysed for users with three different path lengths, 1, 2 and 3, for each of the $N_p^{overlapping}$, $N_p^{cross-region}$ and $N_p^{single-network}$ pairs of participants - Tables B.2, B.3 and B.4. Results show that, for PL of 1, trust metrics for $N_p^{overlapping}$ pairs of participants always improve from one of the individual networks, ePrints or WAIS in all cases, because fusion of two values always results in between the two extreme values available, and in 39% of the cases it results in improvement on both the individual networks. This percentage, however, decreases for $N_p^{cross-region}$ and $N_p^{single-network}$ users. For $N_p^{cross-region}$ users, trust improvement exists in only 16% of the cases as in rest of the cases it deteriorates the trust metrics extracted from individual networks. For $N_p^{single-network}$ pairs of users, in 33% of the cases, trust metrics in MuDi are improved from those obtained from individual networks. For PL of 2, trust between 25% of the $N_p^{overlapping}$ pairs of participants is improved over both the networks, all of the times for $N_p^{cross-region}$ users and 33% of the times for $N_p^{single-network}$ users. For PL of 3, trust from individual network ePrints is improved in 33% of the cases for $N_p^{overlapping}$ users, 26% of the users pairs for $N_p^{cross-region}$ users and all of the times for user pairs from ePrints network.

From these results, it can be said that for $N_p^{overlapping}$ pairs, although the percentage of improvement due to consolidation of MuDi network is high, there is uniformity, with users belonging to each of the path length sharing approximately same percentage of improvement, comparatively better than the other two categories $N_p^{cross-region}$ and $N_p^{single-network}$ pairs. While in the case of $N_p^{cross-region}$ and $N_p^{single-network}$ users, there is 100% improvement for PL of 2 and 3 respectively, with comparatively less improvement in the other two cases. Also the *mean* of the differences of system and survey readings for $N_p^{overlapping}$ and $N_p^{single-network}$ users is far less than those to $N_p^{cross-region}$ users, which shows that there is a high magnitude of error for *cross-region* users that results in deteriorating consolidated trust metrics.

Figure 6.5: The number of overlapping nodes (PO), overlapping ties (TO) (in formation PO/TO) in different portions of consolidated networks are given. Angular lined areas represents ePrints and WAIS networks respectively while vertical lined area shows the overlapping region.

Another aspect that can impact the accuracy of aggregated trust metrics is an unusual overlap between the ePrints and WAIS networks when compared with networks generated in the simulation study. The values of Participant Overlap (PO) and Tie Overlap (TO) in the simulation were uniform across both networks, with the same number of participants assumed in both. But in this study due to different sizes of networks, PO and TO are distinct when analysed with reference to each networks. For the WAIS network, PO and TO are 51% and 78% respectively while they stand at 2% and 1.4% in respect of the ePrints network (see Figure 6.5). This means that these networks are not homogeneously overlapping. In this context, it is quite challenging if the results from this experiment are to be framed in terms of the simulation results. The idea of maintaining the integrity of trust from individual networks while consolidating multiple social networks, in the simulation, was based on the assumption that it would improve the quality of trust from individual networks to that perceived by users about each other in real life. If the results from the real world experiment are to be analysed in that context, then it seems that although simulation experiment has analysed it for varying PO and TO, it still misses the scenario generated in this experiment. So the earlier assumption of relating quality of trust value with the preservance of integrity, although appears to be a valid argument when analysed for a series of networks in the simulation study, in the real world, where the relative size of networks can be very different, only stands true for $N_p^{overlapping}$ users; and for $N_p^{cross-region}$ and $N_p^{single-network}$ users, it is false. This might well be due to different qualities of trust in individual networks, but to isolate the reasons behind this unexpected outcome demands even more exhaustive tests with a range of variable overlaps between networks to draw any conclusion about its impact on trust in real life scenarios.

The issue of varied trust value perception across multiple social networks lies upon two principles. First one was to ensure that trust properties from each of the individual networks are preserved when they are consolidated. This meant that if data from either of the networks is to be normalised with respect to the other then it would result in distorting the essence of the trust from one of the networks. It would either distort or inflate trust values which are not supposed to happen if we assume each network carrying trust in its own domain. Second principle was to acknowledge the fact that there exist trust outliers in the each of the individual networks. Distribution of these outliers should be adjusted keeping in mind the properties of that specific network so as to normalise them for comparing with data from other networks. Cross-normalisation may result in affecting these outliers further which was the loss of information in our scenario. So the assumption about normalisation of trust data from multiple networks with respect to their individual networks was the right choice due to the contextual nature of trust that exit in different networks. It cannot be affirmed whether the resultant corresponding values from multiple networks would be semantically similar (i.e. 0.6 from Facebook in comparison with 0.8 from Twitter), but this would provide data for aggregation that is normalised in its domain network.

6.5 MuDiExperts - Trust-Aware Expert Recommendation over MUltiple DIstributed (MuDi) Networks

One of the potential benefits of consolidating MuDi networks is to improve existing expert recommendation mechanisms, by presenting the better options of experts compared to individual social networks. Existing 'expert finding' systems consider the research profile along with the number of directly related experts as the metric to classify a global list of experts in any domain (Zhang et al., 2007; Crowder et al., 2002a). By contrast, our hypothesis is that trust-aware personalised expert recommendations using consolidated MuDi networks from different domains generates options of experts that better match the expectations of users than individual networks. This is because it not only combines information about users from multiple networks but includes more experts who are not part of individual networks.

6.5.1 Experiment Design

To test this hypothesis, a randomly selected research interest of each *rating* participant was taken into account by analysing the titles of their articles published in ePrints and projects in WAIS (based on their presence in either or both networks). Based on that information, a list of experts from both the networks related to that field was presented, and participants were requested to select their preferred list of experts.

6.5.1.1 Sample Research Area Selection

When users logged-in to the survey, the titles of the ePrints publications and WAIS projects of the *rating* participant (based on the presence of the user in either or both the networks) were compared with the set of research areas of people working in WAIS research group, to find the research area of the participating user. There were eight areas selected for people to answer their query about expert recommendation; *semantic web*, *social networks*, *e-learning*, *agents*, *trust*, *multimedia*, *web services* and *accessibility*. This filtering was undertaken to select the area familiar to the logged-in user because as it would be difficult for the participant to recommend experts about the unknown area. If more than one research areas were returned, then one area was selected from those present by a uniformly random process.

SPARQL ASK Query 19 Selecting research area of the *rating* participant from the Eprints co-authorship network

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX dct:<http://purl.org/dc/terms/>

ASK
FROM NAMED <http://eprints.soton.ac.uk>
WHERE {
GRAPH <http://eprints.soton.ac.uk> {
        ?epuri1 dct:title ?title.
        ?epuri1 dct:creator ?presonuri.
        FILTER (str(?personuri)='"""+eprintsURI+"""').
        FILTER (regex(str(?title),'semantic web',"i")
        || regex(str(?title),'linked data',"i")) }
}
```

SPARQL ASK Query 20 Selecting research area of the *rating* participant from the WAIS projects collaboration network

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX dct:<http://purl.org/dc/terms/>
PREFIX akt: <http://www.aktors.org/ontology/portal#>

ASK
FROM NAMED <http://wais.soton.ac.uk/projects>
WHERE {
GRAPH <http://wais.soton.ac.uk/projects> {
        ?personuri ecs:memberOf ?projecturi1.
        ?projecturi1 dct:title ?title.
        ?akturi owl:sameAs ?personuri.
        FILTER (str(?personuri)='"""+waisURI+"""').
        FILTER (regex(str(?title),'semantic web',"i")
        || regex(str(?title),'linked data',"i")) }
}
```

SPARQL ASK queries 19 and 20 show the sample queries sent to the SPARQL endpoint of the Sesame triplestore for finding if the *rating* participant works in the *semantic web* research area. Two presented ASK queries were sent to the ePrints and WAIS projects networks, respectively, with *eprintsURI* representing URI of the participant from the ePrints publication network and *waisURI* as the URI from WAIS projects collaboration networks, based on presence in these networks. It returns the Boolean answer *Yes/No*, based on whether the term *semantic web* exists in the titles of their research articles or the projects titles of which the *rating* participant was part of, meaning that the logged-in participant was working in that area and that the experts list relating to that area could be presented for recommendation.

6.5.1.2 Experts List Presentation

Experts in the research area were selected in the same way as the selection of the research area, by analysing their publication/project titles, but there was a slight difference. A person in our case was considered to be an expert of some area if he or she had a substantial number of publications, for the initial testing selected to be three. This meant a list of experts for any research area contained those people with three or more publications in that field. Participants had to select their preferred list of experts, bearing in mind the priority to work and interact with them if they ever had the chance.

SPARQL SELECT Query 21 Selecting Participants for Rating

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX dct:<http://purl.org/dc/terms/>

SELECT DISTINCT ?name
FROM NAMED <http://eprints.soton.ac.uk>
WHERE {
GRAPH <http://eprints.soton.ac.uk> {
        ?epuri1 dct:title ?title.
        ?epuri1 dct:creator ?creator1.
        ?creator1 foaf:name ?name.
        FILTER (regex(str(?title),'semantic web',"i")
        || regex(str(?title),'linked data',"i")) }
}
GROUP BY ?name
HAVING(COUNT(?epuri1) > 3)
```

The set of SPARQL SELECT queries 21 and 22 extracted *?names* of the experts with more than three publications in the area of '*semantic web*'. FILTER statement included in both the queries ensured that the strings *semantic web* and related term *linked data* were in the titles of the research articles and projects, from ePrints and WAIS projects networks respectively. The HAVING clause at line 10 (in SPARQL SELECT Query 21) and line 12 (in SPARQL SELECT Query 22) counted and ensured those researchers were returned who had more than three articles/ projects in the area of semantic web/ linked data.

SPARQL SELECT Query 22 Selecting Participants for Rating

```
1: PREFIX foaf:<http://xmlns.com/foaf/0.1/>
2: PREFIX dct:<http://purl.org/dc/terms/>
3: PREFIX akt: <http://www.aktors.org/ontology/portal#>
4: PREFIX owl: <http://www.w3.org/2002/07/owl#>
5:
6: SELECT DISTINCT ?name
7: FROM NAMED <http://wais.soton.ac.uk/projects>
8: WHERE {
9: GRAPH <http://wais.soton.ac.uk/projects> {
        ?personuri ecs:memberOf ?projecturi1.
        ?projecturi1 dct:title ?title.
        ?akturi owl:sameAs ?personuri.
        ?akturi akt:full-name ?name.
        FILTER (regex(str(?title),'semantic web',"i")
        || regex(str(?title),'linked data',"i")) }
10: }
11: GROUP BY ?name
12: HAVING(COUNT(?projecturi1) > 3)
```

6.5.1.3 Application Interface

The user interface for this experiment was a two-list architecture with right and left buttons for moving data across lists (shown in Figure 6.6). The left-hand list was populated with the names of potential experts in the selected research area of the logged-in user, while the right-hand list was initially empty, being populated when users selected the expert of their choice from the left-hand side and transferred the name to the right-hand side. Participants could transfer all the names from the alphabetically arranged list on the left to the right, and at least one expert needed to be selected to be a part of this study; if there were unfamiliar people or experts, one did not want to include them in the list, so they remained on the left. Relevant error or information messages popped up at every stage of the process to make it easy for the *rating* participant.

This portion of the survey page was flexible, designed using *JavaScript* and *JQuery* to keep the interface user-friendly and easy to interact with. When the user had finished the survey and was ready to submit, it provided a single button to submit the survey. The last page of the application thanked the user and gave the option to return to the login page if they wanted to participate again.

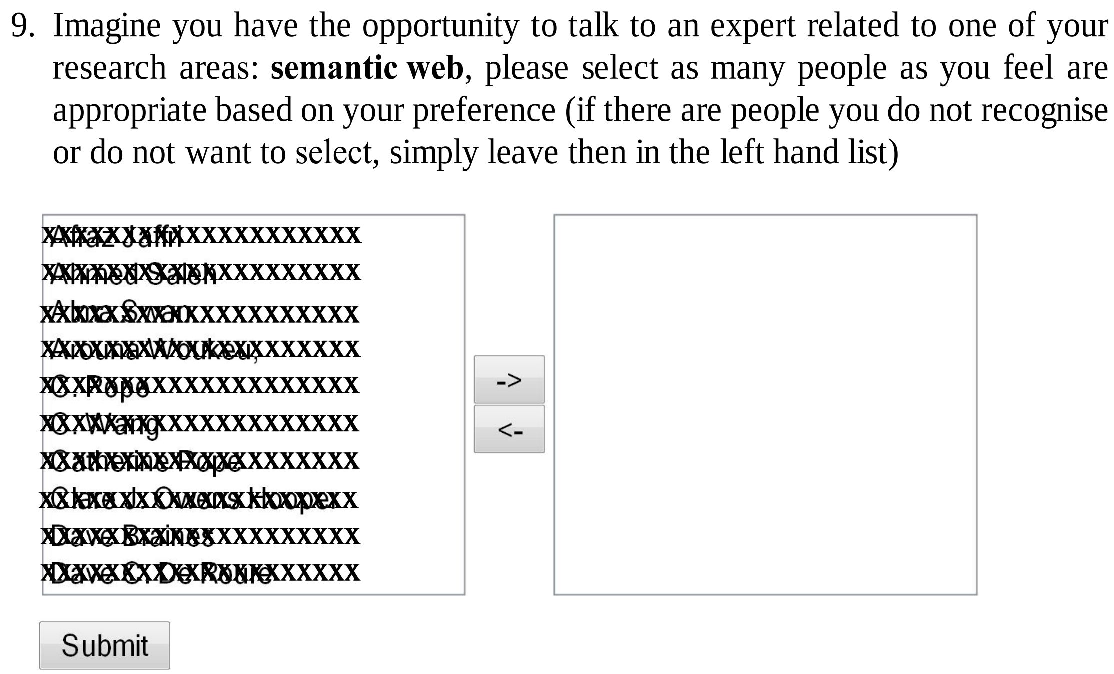


Figure 6.6: Survey application interface for expert recommendation in the area of *semantic web*. Names in the expert list are redacted for anonymisation.

6.5.2 Results and Analysis

A total of 23 participants participated in the expert recommendation portion of this survey; three of the initial group had refused to participate in this part of the survey. Eight participants answered the expert recommendation question about each of the '*semantic web*' and '*social networks*' research areas; four recommended experts related to '*e-learning*', two about '*multimedia*' and one about '*agents*'. As each of the participants was free to select a number of experts, the compiled expert list contained variable number of experts from different *rating* participants who took part in this survey.

Table 6.18: Breakdown of expert recommendation ratings corresponding to different research areas selected.

Research Area	Number of participants
Semantic Web	8
Social Networks	8
E-learning	4
Multimedia	2
Agents	1

6.5.2.1 Expert Recommendation Data Description

The list of experts recommended by *rating* participants (*RPs*) corresponding to different research areas is presented in Table B.1. Each row represents a record corresponding to one of the *RP*, and there are four lists of experts for each *RP*. The *survey expert list* is the one obtained from the survey experiment and was provided directly by those participating in the survey, $TAEL_{ePrints}$ is the *trust-aware expert list* generated from the ePrints network, $TAEL_{wais}$ is the *trust-aware expert list* generated from the WAIS network and $TAEL_{mudi}$ is the one extracted from a consolidated version of ePrints and WAIS. The numeric lists from each of the networks are the encoded version of the experts recommended by the survey participants.

Corresponding to the number of experts recommended by each of the *RP* in the *survey expert list*, the system extracted an equivalent number of top trusted experts from the potential list of experts provided to each *RP* for recommendation (left-hand side of Figure 6.6). It took the *RP* as an ego-node and evaluated trust for all potential experts, selecting the top trusted experts' equivalent of the number of experts recommended by *RP* in the survey experiment. This procedure was repeated for all the three networks - ePrints, WAIS and MuDi - which generated three lists: $TAEL_{ePrints}$, $TAEL_{wais}$ and $TAEL_{mudi}$ respectively, as shown in Table B.1.

Taking the *survey expert list* as a reference, numeric digits in all the three lists represent the survey experts which exist in the list provided by the corresponding network. If there is no list available, it means that the *RP* is not part of that network, hence expert list cannot be extracted, while the digit 0 means none of the trusted experts related to the corresponding *RP* from that network matches with the survey list, (for example *RP5* in the *social networks experts list*).

Comparison of the survey-generated list with each of the system-generated lists helped us to find whether the expert recommendation over MuDi matches the user-provided list of experts better than those from individual networks.

6.5.2.2 Expert Recommendation Data Analysis

The expert recommendation data provided by the *rating* participants (*RPs*) was analysed to see whether the consolidation of MuDi social networks generates trust-aware list of experts that matches to those collected from the *RPs* using an expert recommendation survey better than those from individual networks.

Analysis of this data calculates Jaccard coefficient[4] for evaluating similarity between expert lists extracted from each of the ePrints, WAIS and MuDi with respect to the one

[4] http://en.wikipedia.org/wiki/Jaccard_index

obtained from the survey. The Jaccard coefficient between two sample sets A and B (0 $\leq$ J(A, B) $\leq$ 1) can be calculated using Equation 6.4:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{6.4}$$

Interpreting in terms of survey and system readings J(A, B) can be rewritten as J(Survey, ePrints) for Jaccard coefficient between survey and ePrints readings, J(Survey, WAIS) for Jaccard cofficient between survey and WAIS readings and J(Survey, MuDi) for Jaccard coefficient between survey and system readings from consolidated version of MuDi social networks.

$$J(Survey, ePrints) = \frac{|Survey \cap ePrints|}{|Survey \cup ePrints|} \tag{6.5}$$

$$J(Survey, WAIS) = \frac{|Survey \cap WAIS|}{|Survey \cup WAIS|} \tag{6.6}$$

$$J(Survey, MuDi) = \frac{|Survey \cap MuDi|}{|Survey \cup MuDi|} \tag{6.7}$$

If the following hypothesis is '*true*', it means that the trust-aware expert recommendation from consolidated MuDi networks gives improved results to those calculated from individual networks.

Hypothesis: If J(Survey, MuDi) $\geq$ J(Survey, ePrints) and J(Survey, MuDi) $\geq$ J(Survey, WAIS), this shows that the expert recommendation from MuDi networks is better or similar to individual networks ePrints and WAIS

Null Hypothesis: If J(Survey, MuDi) $<$ J(Survey, ePrints) or J(Survey, MuDi) $<$ J(Survey, WAIS), this shows that the expert recommendation from individual network ePrints or WAIS performs better than MuDi networks.

Table 6.19: Jaccard similarity cofficients between expert recommendations from survey and Trust-Aware Expert List (*TAEL*) from ePrints, WAIS and consolidated MuDi networks for different *rating* participants (*RPs*)

Rating Participant (RP)	J(Survey, ePrints)	J(Survey, WAIS)	J(Survey, MuDi)
Semantic Web Experts			
RP1	0.33	-	0.33
RP2	0.6	-	0.6

Continued on next page

Table 6.19 – continued from previous page

Rating Participant (RP)	J(Survey, ePrints)	J(Survey, WAIS)	J(Survey, MuDi)
RP3	0.50	0.13	0.50
RP4	0.47	0.05	0.57
RP5	0.33	0.07	0.45
RP6	0.20	-	0.20
RP7	0.33	0.10	0.33
RP8	0.11	0.11	0.11
Social Network Experts			
RP1	0.11	0.25	0.11
RP2	0.2	-	0.5
RP3	0.25	-	0.25
RP4	0.20	-	0.20
RP5	0	0	0
RP6	0.25	0.25	0.43
RP7	0.25	-	0.25
RP8	0.33	-	0.33
E-Learning			
RP1	0.50	0.08	0.50
RP2	0.40	-	0.53
RP3	0.54	-	0.54
RP4	0.50	0.13	0.64
Multimedia			
RP1	0.57	-	0.57
RP2	0.45	0.6	0.33
Agents			
RP1	0.14	-	0.23

Table 6.19 presents the values of Jaccard similarity coefficients for different sets of expert recommendations data generated by the survey and the system. Missing data from certain network is due to the *RP* not being present in that network, while 0 shows dissimilarity between the survey and the system-generated data.

Analysis of the results between ePrints and consolidated MuDi networks revealed that in 30% of cases, J(Survey, MuDi) > J(Survey, ePrints), which shows the percentage of times MuDi social networks perform better than ePrints. In 4% of the cases, J(Survey, ePrints) > J(Survey, MuDi), meaning that ePrints gives better results, while results from both ePrints and MuDi are similar in 66% of the cases, that is, J(Survey, MuDi) = J(Survey, ePrints).

Analysing it between WAIS and MuDi networks showed that J(Survey, MuDi) > J(Survey, WAIS) stands true in 82% of the cases, in 9% of the times, J(Survey, WAIS) > J(Survey, MuDi) meaning that WAIS performs better than MuDi networks, while in 9% cases, J(Survey, WAIS) = J(Survey, MuDi), that is, both the networks perform similar.

Looking at the results with respect to the hypothesis proves that J(Survey, MuDi) at least equals J(Survey, ePrints) in 96% of the times, being worse in 4% of the times. On the other hand, when compared between WAIS and MuDi, hypothesis stands true (meaning J(Survey, MuDi) at least equals J(Survey, ePrints) in 91% of the cases and false in 9% of the cases. Apparently the results implicates that WAIS performs comparatively better than ePrints when analysed for this hypothesis, but further discussion about what it tells actually is presented in Section 6.5.2.3.

6.5.2.3 Discussion

The consolidation of MuDi professional networks for trust-aware expert recommendation provides an opportunity to exploit participant's connections from multiple social networks to explore better options of experts. Existing expert finding systems like ArnetMiner[5] Tang et al. (2008) present experts, based on their expertise across different publication repositories all over the web, but these systems are inadequate in the sense that most of the users who are seeking experts have never seen or met those experts in their life. As a result they hesitate to interact with or approach them due to having a low level of trust in them. This makes 'expert finding systems dysfunctional and their purpose becomes merely to provide information rather than to meet the needs of users.

The consolidation of MuDi networks and one such prototype, *MuDiExperts*, tested in this thesis implements a trust-aware expert recommendation mechanism that creates an opportunity to generate a personally trusted list of experts spanning multiple social networks. These are the experts whom either the querying person knows directly due to previous work experience with them or through working with someone else whom the querying person knows directly or indirectly. In a way it uses a trust propagation mechanism to discover links based on either direct trust or transitive trust between the person searching for an expert and the potential expert. As a result, it would end up presenting experts from the same research group, same research institute as the querying person or from somewhere the known person is working in.

The results of the *MuDiExperts*' prototype tested in this work support our argument, because expert lists provided by MuDi networks were either similar or closer to the survey lists than those given by the individual networks. If analysed with respect to each of the individual networks, MuDi performed better than ePrints in 30% of the cases, while in 82% of the cases than WAIS. Similarly ePrints performed better than MuDi

[5] http://arnetminer.org/

networks in 4% of the cases compared to 9% for WAIS. In 66% of the cases ePrints generated results similar to MuDi compared to 9% of the times in WAIS. Although WAIS apparently showed better results than ePrints in terms of improvement, this relative improvement of WAIS (due to improved results in 9% of the cases) happened for only certain cases, because in 82% of cases it ended up performing worse than MuDi (compared to ePrints which performed worse in only 30% of the cases) apart from 9% of cases when it performed similar. This indicates that the ePrints was a stable and substantive network than WAIS projects network in this experiment. However, it remains an open question whether combining two relatively accurate networks might result in an even more accurate MuDi network.

6.6 Conclusions

This chapter presented the analysis of consolidating real world MuDi professional networks by selecting a pair of networks, ePrints co-authorship network and the WAIS projects collaboration network in the University of Southampton domain. To judge the benefits of consolidation and the accuracy of the aggregated trust metrics, a trust survey was run in the University domain that collects real life proxy professional trust between researchers working in the same environment. Analysis of the results revealed that the consolidation of MuDi networks performed better for overlapping users than both the ePrints and WAIS networks. For cross-region participants with just one of the users in more than one network and the other user a part of one, it deteriorates the results, while it maintained the same results for pairs that are part of only one of the networks.

The proposed idea of consolidating MuDi networks was then used for expert recommendations and a system, MuDiExperts, was tested that extracts trust-aware experts lists from multiple social networks. Part of the trust survey used for recording real life trust metrics also collected experts lists from participating users relating to one of their research area and the results from individual and consolidated networks were compared with these survey results. Analysis of the results revealed that MuDi performed better than ePrints in 30% of the cases, similar in 66% of cases, while deteriorated results in 4% of the cases. On the other hand, MuDi performed better than WAIS in 82% of the times, deteriorated in 9% of the time, while generated results similar to WAIS in 9% of the cases.

The next chapter relates the results obtained in Chapters 4 and 5 with the set of hypotheses specified in Section 1.3 to discuss the overall objectives achieved in this thesis. Furthermore, it explicitly states the contributions of this thesis in the areas of trust, semantic web and social networks.

Chapter 7

Conclusions

7.1 Summary

Use of digital networks is increasing and the presence of users in multiple networks is a great opportunity to make trust metrics on the web more sophisticated by incorporating a variety of information. The idea is to draw interaction information between participants from a variety of networks that represent multi-context trust, to explore whether the consolidated trust captures real life trust better than individual networks.

This thesis moves one step towards that end and paves the way by proposing a semantic web framework, MuDiTCF, that consolidates multiple social networks and makes trust computations over them. Networks are interlinked by identifying participants who exist in both the social networks using the co-reference resolution mechanism, and trust measurements available between them are aggregated using data fusion techniques. Trust between those participants having no direct connections is derived using the trust transitivity principle that considers the decay of trust over paths.

Two sets of experiments were run. The simulation experiment was to select the data fusion technique that best preserved the integrity of trust from individual social networks. Pairs of networks with varying percentages of Participant Overlap (PO) and Tie Overlap (TO) were generated and consolidated to analyse this trust aggregation property. Naive techniques Sum (S), Weighted Average (WA) and Induced Ordered Weighted Averaging (IOWA) distorted the trust from individual social networks; only Weighted Ordered Weighted Averaging (WOWA) turned out to respect the integrity of trust for different values of PO and TO. The real-world experiment used the recommendation of the simulation finding that the WOWA technique was the best to consolidate a pair of professional social networks: ePrints co-authorship network and WAIS projects collaboration network. Both were extracted from the University of Southampton domain. Data were analysed for overlapping, cross-region and single-network pairs of participants. In addition to the abovementioned experiments, an expert recommendation application was also designed, that tested the proposed framework in context of real world decision making of the users. The participants of the application were asked to recommend experts from the provided lists of experts related to one of their research areas. These lists were then compared with those generated from individual and MuDi networks taking each of the participants as an ego-node. Data were analysed to asses which of the expert lists better matched with the user provided lists of experts.

The analysis completed in this work partially validates the concept of consolidating MuDi networks. Out of the three types of participant pairs analysed from the consolidated pair of networks, aggregated metrics turned out to be better for overlapping pairs of participants ($p \leq 0.05$), while the hypothetical claim was disproved for cross-region and single-network users ($p > 0.05$). Results from the expert recommendation experiment proved the consolidation of MuDi networks as a productive approach. In 30% of the cases MuDi recommended experts better than ePrints, and 82% of the times better than WAIS. While 4% of the times with respect to ePrints and 9% corresponding to WAIS, it deteriorated the results. In rest of the cases, (that is, 66% for ePrints and 9% for WAIS), it generated similar results.

7.2 Hypothesis Review

The work completed in this thesis tests all three claims of the hypothesis and this section reviews the findings corresponding to each.

H1 *Semantic technologies allow us to uniformly model and annotate trust data from MuDi social networks for making trust computations over heterogeneous resources.*

This hypothesis was tested in Chapter 3 and the following results were found.

Implementation - The proposed idea of using semantic technologies for consolidating multiple social networks and making trust computations over heterogeneous networks was implemented using a semantic web framework. It interlinked different social networks by identifying participants that exist in multiple social networks using the concept of co-reference resolution. Trust metrics between participants from networks using different ontologies were incorporated in a proposed ontology, thus presenting it in a uniform representation. The proposed trust ontology extended one that was already developed and included classes and properties specifically needed for defining trust over multiple social networks.

Evaluation - To evaluate the proposed framework, a set of simulation and real world experiments were run on top of the framework for a pair of social networks. Both individual and consolidated versions of these networks were stored as named graphs in Sesame triplestore and SPARQL queries were executed to annotate and query semantic data from these networks. The successful execution of these queries justified the affordances of using semantic technologies for evaluating trust over heterogeneous social networks.

Results -

1. The co-reference resolution successfully classified participants available in both the networks into two sets. The co-referred set had participants who exist in both networks, while the non-co-referred set contained the rest of the participants from the individual networks. A separate namespace was used to generate URIs for the co-referred participants in the consolidated version of the networks. This was represented as a separate named graph in the triplestore and *owl:sameAs* predicate was used to refer the newly created URI to both of the individual network URIs.

2. The proposed version of the MuDi trust ontology recorded all the properties needed to represent consolidated trust. For example, as explained in Section 3.6.2, it annotated updated strength of trust tie (i.e. *trust:has_processedValue* $\longrightarrow$ *0.8*), updated length of trust path (i.e. *trust:has_pathLength* $\longrightarrow$ *1*), added trust aggregation technique used (i.e. *trust:has_aggregationTechnique* $\longrightarrow$ '*weighted ordered weighted averaging*') and so on. These properties were then compared with those obtained from individual networks for evaluating the performance of consolidation in the next phase of the work.

H2 *Data fusion techniques allow us to aggregate trust metrics from MuDi social networks and respect the integrity of trust from individual networks, while opening up many additional trust paths.*

This hypothesis claim was tested in Chapter 5 using a simulation and the following were the findings of this experiment.

Implementation - When consolidating multiple social networks, two types of trust data emerged between participants present in multiple networks. Either trust metrics between them were available from both the networks due to them being co-referred users, present in multiple networks, or from a single network due to being member of just one of the networks. The data fusion technique aimed to respect the integrity of trust available from individual social networks while aggregating trust information between pairs of participants from multiple social networks. This means that it should neither inflate aggregated value if data is available from both the networks nor dampen down if available from one of the networks.

Evaluation - To evaluate the performance of different data fusion techniques for trust aggregation, a simulation experiment was designed that generated a set of networks with varied percentages of Participant Overlap (PO) and Tie Overlap (TO). An experiment was run for four different data fusion techniques: Sum (S); Weighted Average (WA); Weighted Ordered Weighted Averaging (WOWA); and Induced Ordered Weighted Averaging (IOWA). Two sets of results were compiled for each data fusion technique. The average strength of trust ties (TS) metric was used to analyse whether the integrity of trust from individual networks is respected in consolidated version of the networks, and average length of trust paths (TL) metric was to evaluate whether the consolidation of networks resulted in creating new trust paths.

Results -

1. The TS metric calculated for WOWA justified the hypothesis claim and respected the integrity of trust under varied values of *PO* and *TO*. At low *PO* and *TO*, it successfully managed to differentiate *absence of trust* from *distrust* due to missing trust information between pairs of participants from one of the network, and at high *PO* and *TO* it restricted the metric from inflating the trust. The T-Test applied for determining statistical significance also proved the claim as the p-value for trust calculations between WOWA and other data fusion techniques was less than 0.05, that is, $p < 0.01$. The rest of the data fusion techniques turned out to be naive for trust aggregation operation as they were unable to satisfy the hypothesis for varying *PO* and *TO*.

2. The calculated TL metric justified the second part of the hypothesis (about opening up additional trust paths) as its value decreased in the consolidated version of MuDi networks for all data fusion techniques. At low *PO*, the decrease was not highly significant as a limited *PO* resulted in a bottleneck in the consolidated network. However, when *PO* increased, it opened up new trust paths that generated trust metrics with shorter trust paths. When tested for statistical significance by applying T-Test, WOWA turned out to be best as its TL metric was significantly better among all the data fusion techniques represented ($p < 0.01$).

H3 *Trust metrics generated over MuDi social networks increase accuracy of trust over individual networks in terms of more accurately caputring how users perceive one another in real life.*

Chapter 6 tested this hypothesis claim by running a real world experiment and the following results were found.

Implementation - Using the proposed semantic web framework and based on the recommendation from the simulation experiment, WOWA was used to aggregate trust between pairs of participants in a pair of professional social networks: ePrints and WAIS. Co-reference resolution revealed that the WAIS network was approximately a subset of ePrints, meaning there were significant *PO* and *TO* with respect to WAIS but not with respect to ePrints. This was because the number of participants and ties in WAIS were far fewer than in ePrints.

Evaluation - To test the hypothetical claim that trust metrics generated from consolidated networks are more accurate in terms of closeness to the user perception, trust between users pairs was evaluated for three different types of participant pairs. The first type was overlapping participants in which both members of the pair existed in both networks; the second was *cross-region* pairs in which one of the members was in both the networks; and in third type, *single-network* pairs, both members were in a single network. Besides this, a survey was run that asked participants proxy trust questions in an attempt to ascertain the professional trust that participant pairs have of each other in real life. The trust metrics from individual and consolidated networks were then tested to see which set of trust values correlated most closely to the proxy trust questions. To further analyse whether the claim of generating better trust metrics using consolidated networks works in real world scenarios, an expert recommendation application was designed that asked participants to select a list of preferred experts from the list of available relating to one their research areas. This was to evaluate whether the expert recommendations from MuDi networks are more similar to those provided by users than individual networks.

Results -

1. The claim of this hypothesis for increased accuracy of trust was justified for overlapping participants. Statistical significance was calculated by applying T-Test and $p \leq 0.05$ of system trust metrics available from MuDi networks and survey trust metrics, showing that the consolidated network for overlapping pairs of participants captures real life trust better than individual networks.

2. For *cross-region* users the hypothesis was proven to be wrong as it deteriorated the trust value from individual networks in the sense that it took the values further from what had been given by the participants of the survey. The p-value was although less than 0.05 but when the means of the differences between system and survey readings were analysed, the mean of absolute difference from consolidated network was high compared to the mean of the difference between system and survey readings from individual networks. This proved that, in this case, the significant difference between system and survey readings was not because consolidation of MuDi networks did not bring it close enough to decrease the gap between system and survey readings. Instead it took it further away, increasing the magnitude of difference between system and survey readings.

3. For *single-network* users, the hypothesis was again proved wrong, but here it did not result in deteriorating trust values, but rather maintained the same result as individual networks. It was proven using the T-Test, when the p-value was greater than 0.05 (i.e. $p > 0.05$).

4. The designed expert recommendation application justified the validity of the hypothesis in a real world scenario. The calculated Jaccard coefficient for evaluating

similarity between survey list and lists provided by individual and consolidated MuDi networks proved that MuDi networks improved expert recommendation results in 30% of cases compared to ePrints while keeping it same in 66% of the cases. In only 4% of the cases it deteriorated the results. For WAIS network, MuDi improved results in 82% of the cases, generated similar results in 9% of cases, and deteriorated results in 9% of the cases.

7.3 Contributions

This thesis contributed the following to the existing literature relating to trust and the semantic web.

1. The affordances of semantic web technologies for making trust computations over heterogeneous social networks are successfully displayed by testing a semantic web framework that allows to consolidate multiple social networks.
2. The usability of data fusion techniques for aggregating multiple trust metrics while respecting the integrity of trust from individual networks is proven.
3. The idea of consolidating multiple social networks for trust-related decision making is now validated and proven to be a successful approach for users who share at least one common network.

Beside the abovementioned contributions, the simulation portion of this work was also published in the ASE Human Journal Imran et al. (2012).

7.4 Limitations

The work completed in this thesis also has some constraints that need to be mentioned.

1. One of the limitations is about using a pair of networks (ePrints and WAIS) for real world experiment in this work. Although the work analyses the trust metrics between different pairs of participants (overlapping, cross-region and single-network), the comparative dissimilarity between size and/or overlap of networks may be problematic. The simulation portion of the work, however, tries to mitigate this weakness by analysing it over a range of networks with different percentages of overlap. Still the overlap of the real networks is out of proportion to the those tested in the simulation experiment (with 50%PO of the overlap in respect of the WAIS and 2%PO of the overlap in respect of ePrints) that makes it a unique case.

2. This work respects the integrity of trust values from individual networks for all the participants uniformly, but in practice network members may have individual priorities of networks. Some users can well specify certain network to be more trustworthy which others in the network believed to be otherwise. In such situations, the trustworthiness/reliability of the individual networks can be taken as an input from users before measuring consolidated trust metrics over multiple social networks.

3. The semantic techniques and especially co-reference resolution used in this work are relatively simple, especially if we analyse with respect to the varied types of ontologies and semantic data available on the web. In this perspective, either a co-reference resolution repository that runs independently and carries all the resolved URIs from the web should be established, or an existing one could be used for this purpose, for example, *www.sameas.org*.

4. The system-generated numerical trust values are compared with those available from likert scale collected using a survey. This may be problematic as it arises question about how well can a system-generated continuous value map on a discrete set of values collected using likert scale. Although there is no exact answer about the accuracy of such comparison but the purpose of this study was not to select the best comparison out of the different available. It aimed to conduct a relative comparison between system and survey trust metrics. So the outcome of the analysis is based on the same set of mapping technique used when comparing trust values from both individual networks and their consolidated version. If using some other technique could change the results then we may expect same effect of change to happen on all the readings uniformly.

7.5 Future Work

Future extensions to this work could test the MuDi consolidation strategy in the context of other scenarios missed in this study.

7.5.1 Using other data fusion techniques

In addition to the WOWA that considers importance of information and source of that information, there are other data fusion techniques that include the confidence of the user and the reasonability of the data, as well (Yager, 2004).

Based on user confidence, the credibility of the source providing the data is evaluated and a decision is made about whether it is acceptable or not. The reasonability of the data ensures that there are no data conflicts and it is measured in comparison with what should be the case. For example, if any professional network rates Tim Berners-Lee not as an expert of the web, this cannot be considered as a reasonable data.

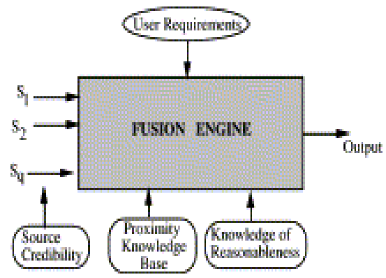


Figure 7.1: Schematic framework of multi-source trust data fusion that considers credibility of data source and reasonability of trut data, (taken from Yager (2004)).

7.5.2 Testing the hypothesis with different trust algorithms

Although this work has partially proved a consolidation strategy as a better approach to trust evaluation, its analysis is based on two trust algorithms that considered decay of trust along paths, if it is to be calculated for distant participants. There are many trust algorithms discussed in the literature that could also be used to test the consolidation approach, for example, algorithms from peer-to-peer networks that punish users for misbehaviour and then takes a long time for a revival of that trust. Unlike the algorithm in this study that calculated a subjective value of trust, this can also test for ecommerce and peer-to-peer trust that considers trust as a global value and uses reputation-based algorithms.

7.5.3 Incorporating other types of trust networks

In the background section, different categories and types of web-based social networks are mentioned that are in use today. These include networks belonging to different categories such as friendship, religious, political, dating and so on. Further, in each category, there are multiple networks such as Facebook and Twitter, both friendship networks, while Linkedin and ePrints are both professional networks.

The existing implementation was limited to professional co-authorship networks in the vicinity of the University of Southampton. It could be extended to the different categories and types of networks described above to see which area best benefits from the consolidation strategy. Furthermore, rather than limiting trust data to co-authorship, even in professional networks trust networks can be generated by incorporating different type of activities, for example those who cite articles of other researchers have a potential trust relationship in the area of research.

7.5.4 Testing the system with higher number of MuDi networks

The current implementation considered only a pair of networks for both simulation and real world network studies. This generated fewer trust aggregation scenarios and as a result made the trust aggregation operation comparatively simple. In future, it could be extended to use a higher number of MuDi networks to test whether the impact on accuracy of trust metrics increases with the number of networks consolidated. By incorporating a variety of networks, it would ensure a more generic and broader perspective of trust than the one calculated on a mere pair of networks.

7.5.5 Incorporating other semantic ontologies

This work deals with only two types of ontologies available from a pair of networks. But data about social networks on the web is available in many other ontologies and schemas. Extension of the proposed framework to include all these ontologies will make it more generic and useful in different scenarios. In future, it can be made to include ontologies representing data from communities using, for example, semantically-interlinked online communities (SIOC).

Another extension to the current implementation could be to set up an independent co-reference resolution mechanism that could more intelligently classify co-referred participants from different social networks on the web. The existing system compares just meta-data (Sleeman and Finin, 2010b), but it could be extended to include supervised machine learning techniques that consult sets of training data to classify co-referred participants. Also, existing repositories holding information about co-referred URIs (for example, www.sameas.org) could be consulted for this purpose.

7.6 Conclusions

This chapter concludes the thesis and presents the summary of the work completed and a set of future research guidelines:

- A semantic web framework was developed that allowed for the consolidation of trust networks, and the calculation of trust metrics over multiple social networks.
- A simulation has shown that in principle networks can be consolidated using the WOWA method maintaining the integrity of tie strength, while still increasing the number of possible trust paths.
- The proposal for generating better trust metrics using consolidated MuDi networks that match real life trust between users was shown to be true for overlapping pairs

of participants, while proven wrong for cross-region and single-network pairs of users.

The section on future work proposes new research dimensions emerging from this study and briefly describes each of them. The existing system could be tested for other data fusion techniques that consider additional trust parameters for validation of data from multiple social networks. It could be run against trust algorithms from peer-to-peer networks and ecommerce that give a global perspective of trust. Furthermore, it could be made to include a higher number and a greater variety of networks to analyse the impact of consolidating multiple networks. The use of machine learning algorithms for co-reference resolution and inclusion of other semantic ontologies is also a fruitful research opportunity.

The hope is that this initiative will encourage trust systems to go beyond their individual networks for decision making. This will improve existing web-based trust systems to make more intelligent trust decisions by incorporating a variety of information about individuals on the web.

Appendix A

Social media matrix

% of users of each particular site who use another particular site (e.g., 29% of Pinterest users also use Twitter)

	Use Twitter	Use Instagram	Use Pinterest	Use LinkedIn	Use Facebook
% of Twitter users who...	N/A	53	34	39	90
% of Instagram users who...	53	N/A	37	30	93
% of Pinterest users who...	29	31	N/A	29	87
% of LinkedIn users who...	31	24	28	N/A	83
% of Facebook users who...	22	23	25	25	N/A

Pew Research Center's Internet Project August Tracking Survey, August 07 –September 16, 2013. Interviews were conducted in English and Spanish and on landline and cell phones.

PEW RESEARCH CENTER

Figure A.1: Pew Research survey results show the percentage of participants overlap (PO) between different MuDi social networks.

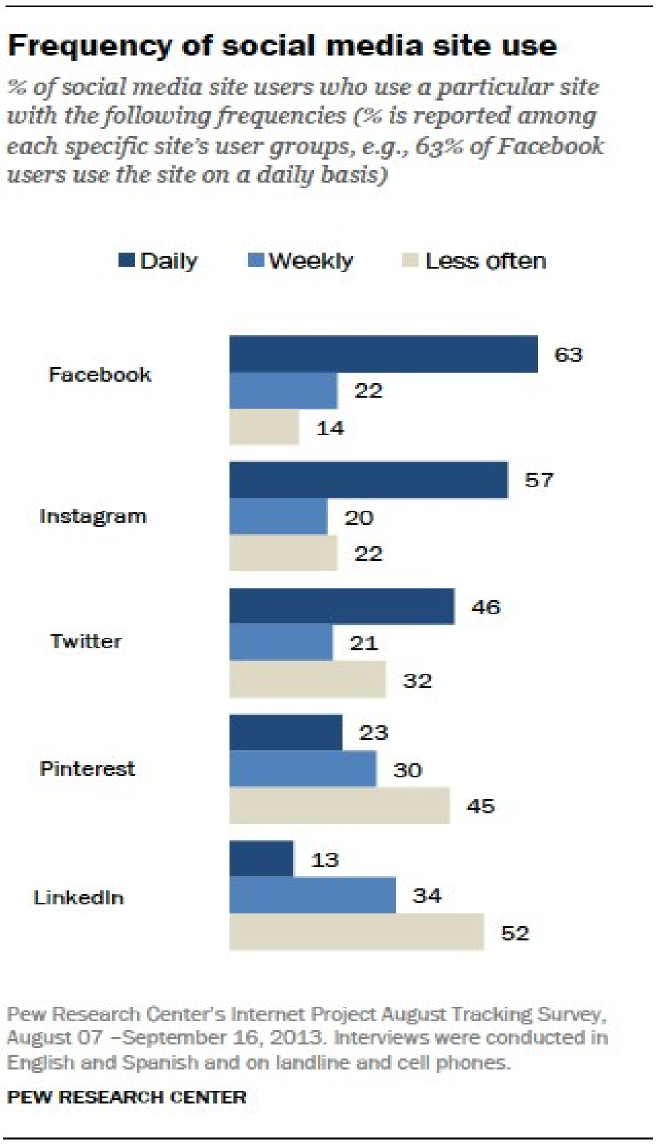


Figure A.2: Pew Research survey results show the percentage frequency of social media site users.

Appendix B

Table B.1: Expert recommendation lists from survey and different professional social networks corresponding to different research areas. $TAEL_{ePrints}$ is the Trust-Aware Expert List extracted from ePrints network, $TAEL_{wais}$ represents Trust-Aware Expert List extracted from WAIS and $TAEL_{mudi}$ is the one extracted from consolidated version of ePrints and WAIS.

Rating Participant (RP)	Survey Expert List	$TAEL_{ePrints}$	$TAEL_{wais}$	$TAEL_{mudi}$
Semantic Web Experts				
RP1	1,2	1,0	-	1,0
RP2	1,2,3,4	1,0,3,4	-	1,0,3,4
RP3	1,2,3,4,5,6,7,8,9	1,2,3,6,7,9	6,9	1,2,3,6,7,9
RP4	1,2,3,4,5,6,7,8,9,10,11	2,4,5,8,9,10,11	11	1,2,4,5,8,9,10,11
RP5	1,2,3,4,5,6,7,8	1,3,6,7	7	1,3,6,7
RP6	1,2,3,4,5,6,7	2,7	-	2,6
RP7	1,2,3,4,5,6	1,4,5,6	6	1,4,6
RP8	1,2,3,4,5	3	4	3
Social Network Experts				
RP1	1,2,3,4,5	1,2	1,5	1,2
RP2	1,2,3	1	-	1,3
RP3	1,2,3,5	2,5	-	2,5
RP4	1,2,3	1	-	1
RP5	1,2,3,4	0	0	0
RP6	1,2,3,4,5	2,5	2,5	2,4,5
RP7	1,2,3,4,5	2,3	-	2,3
RP8	1,2,3,4	1,4	-	1,4
E-Learning				
RP1	1,2,3,4,5,6	1,2,4,5	1	1,2,4,5

Continued on next page

Table B.1 – continued from previous page

Rating Participant (RP)	Survey Expert List	$TAEL_{ePrints}$	$TAEL_{wais}$	$TAEL_{mudi}$
RP2	1,2,3,4,5,6,7,8, 9,10,11,12,13	1,2,4,5,68,12,13	-	1,2,4,5,6,8,11,12, 13
RP3	1,2,3,4,5,6,7,8, 9,10	1,2,3,4,5,6,10	-	1,2,3,4,5,9,10
RP4	1,2,3,4,5,6,7,8, 9	1,2,4,5,7,8	1,9	1,2,4,5,7,8,9
Multimedia				
RP1	1,2,3,4,5,6,7,8, 9,10,11	1,2,4,5,7,8,10,11	-	1,2,3,4,5,7,8,11
RP2	1,2,3,4,5,6,7,8	1,2,3,4,5	1,3,4,5,6,8	1,33,4,5
Agents				
RP1	1,2,3,4,5,6,7,8	6,8	-	4,6,8

Table B.2: Trust ratings from system and survey with each row representing meansurements between a pair of participants, $trust^{eprints}$ reprsents ePrints data, $trust^{wais}$ represents data from WAIS projects network and $trust^{mudi}$ shows data from consolidated version of ePrints and WAIS. Rel and E_M shows relationship and expertise match between participants, PL represents path length between participants in co-authorship network. SP in the Rel represents Supervisor, TM is the abbreviation of Team Member and WC shows those WAIS research group colleagues.

No	System Readings						Survey Readings		Rel	PL	E_M
	Strongest Path Algo			Shortest Path Algo							
	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
1	0.15	0.07	0.14	0.15	0.07	0.14	0.8	0.4	SP	1	0.6
2	0.30	0.06	0.28	0.3	0.06	0.28	0.8	0.2	SP	1	0
3	0.47	0.27	0.45	0.47	0.27	0.45	0.8	0.6	SP	1	0.4
4	0.52	0.43	0.51	0.52	0.43	0.51	0.8	0.8	SP	1	0.8
5	0.75	0.81	0.8	0.75	0.81	0.8	0.8	0.8	TM	1	0.6
6	0.62	0.54	0.61	0.62	0.54	0.61	0.2	0	TM	1	0.2
7	0.54	0.93	0.89	0.54	0.93	0.89	0.8	0.8	TM	1	0.8
8	0.28	0.43	0.42	0.28	0.43	0.42	0.4	0.6	TM	1	0.2
9	0.15	0.27	0.26	0.15	0.27	0.26	0.2	0.4	TM	1	0.2
10	0.30	0.27	0.30	0.30	0.27	0.30	0.4	0.6	TM	1	0.6
11	0.15	0.27	0.26	0.15	0.27	0.26	0.6	0.8	TM	1	0.6
12	0.30	0.27	0.30	0.30	0.27	0.30	0.2	0.4	TM	1	0.2
13	0.55	0.63	0.62	0.55	0.63	0.62	0.8	0.8	WC	1	0
14	0.18	0.27	0.26	0.18	0.27	0.26	0.8	0.4	WC	1	0.6
15	0.14	0.27	0.26	0.14	0.27	0.26	0.4	0.6	WC	1	0.4
16	0.22	0.27	0.27	0.22	0.27	0.27	0.4	0.2	WC	1	0.2

Continued on next page

Table B.2 – continued from previous page

No	System Readings						Survey Readings		Rel	PL	E_M
	Strongest Path Algo			Shortest Path Algo							
	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
17	0.13	0.27	0.26	0.13	0.27	0.26	0.6	0.4	WC	1	0.8
18	0.30	0.27	0.30	0.3	0.27	0.30	0.4	0.6	WC	1	0.2
19	0.15	0.43	0.40	0.15	0.43	0.40	0.2	0.4	WC	1	0.2
20	0.15	0.54	0.50	0.15	0.54	0.50	0.6	0.4	WC	1	0.4
21	0.15	0.07	0.14	0.15	0.07	0.14	0.6	0.6	WC	1	0.4
22	0.24	0.27	0.27	0.24	0.27	0.27	0.6	0.6	WC	1	0.6
23	0.50	0.27	0.48	0.5	0.27	0.48	0.4	0.4	WC	1	0.2
24	0.57	0.15	0.52	0.57	0.15	0.52	0.8	0.6	SP	2	0.6
25	0.15	0.12	0.14	0.15	0.07	0.14	0.2	0.6	TM	2	0.4
26	0.12	0.13	0.13	0.11	0.12	0.09	0	0	WC	2	0
27	0.31	0.17	0.33	0.31	0.17	0.33	0	0	WC	2	0
28	0.15	0.19	0.14	0.15	0.12	0.14	0	0.2	WC	2	0
29	0.23	0.57	0.52	0.19	0.52	0.48	0.4	0.4	WC	2	0.2
30	0.07	0.12	0.12	0.04	0.12	0.07	0	0	WC	2	0
31	0.18	0.17	0.17	0.18	0.17	0.17	0	0	EC	2	0
32	0.1	0.04	0.08	0.1	0.04	0.08	0	0	WC	3	0.2
33	0.17	0.10	0.20	0.06	0.1	0.07	0	0	WC	3	0.2
34	0.14	0.01	0.12	0.11	0.01	0.08	0.2	0.4	WC	3	0.2
35	0.37	0.11	0.37	0.21	0.11	0.22	0	0	WC	3	0
36	0.08	0.03	0.07	0.04	0.02	0.07	0.2	0.2	WC	3	0.4

Continued on next page

Table B.2 – continued from previous page

No	System Readings						Survey Readings		Rel	PL	E_M
	Strongest Path Algo			Shortest Path Algo							
	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{wais}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
37	0.25	0.03	0.25	0.25	0.02	0.22	0.2	0.2	EC	3	0

Table B.3: Trust ratings from system and survey with each row representing meansurements between a pair of participants, $trust^{eprints}$ represents data from ePrints network and $trust^{mudi}$ shows trust metrics from consolidated version of ePrints and WAIS. Rel and E_M shows relationship and expertise match between participants, PL represents path length between participants in co-authorship network. SP in the Rel represents Supervisor, TM is the abbreviation of Team Member and WC shows those WAIS research group colleagues.

No	System Readings				Survey Readings		Rel	HC	E_M
	Strongest Path Algo		Shortest Path Algo						
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
1	0.24	0.22	0.24	0.22	0.8	0.4	SP	1	0.8
2	0.24	0.22	0.24	0.22	0.8	0.8	SP	1	0.8
3	0.15	0.14	0.15	0.14	0.8	0.4	SP	1	0.4
4	0.24	0.22	0.24	0.22	0.8	0	SP	1	0.4
5	0.43	0.39	0.43	0.39	0.8	0.6	SP	1	0.4
6	0.24	0.22	0.24	0.22	0.8	0.8	SP	1	0.4
7	0.24	0.22	0.24	0.22	0.8	0.8	SP	1	0.8
8	0.35	0.32	0.35	0.32	0.8	0.4	SP	1	0.6
9	0.24	0.22	0.24	0.22	0.8	0.6	SP	1	0.2
10	0.15	0.14	0.15	0.14	0.4	0.4	SP	1	0.6
11	0.39	0.35	0.39	0.35	0.8	0.8	SP	1	0.8

Continued on next page

Table B.3 – continued from previous page

No	System Readings				Survey Readings		Rel	HC	E_M
	Strongest Path Algo		Shortest Path Algo						
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
12	0.3	0.27	0.3	0.27	0.8	0.8	SP	1	0.8
13	0.24	0.22	0.24	0.22	0.8	0	SP	1	0.2
14	0.15	0.14	0.15	0.14	0.6	0.4	SP	1	0.6
15	0.15	0.14	0.15	0.14	0.8	0	TM	1	0.4
16	0.3	0.27	0.3	0.27	0.2	0.2	TM	1	0.2
17	0.18	0.16	0.18	0.16	0.6	0.6	TM	1	0.4
18	0.42	0.38	0.42	0.38	0.8	0.6	TM	1	0.6
19	0.15	0.14	0.15	0.14	0.6	0.4	TM	1	0.2
20	0.15	0.14	0.15	0.14	0.6	0.2	TM	1	0.2
21	0.23	0.21	0.23	0.21	0.4	0	WC	1	0.2
22	0.15	0.14	0.15	0.14	0.4	0.4	WC	1	0.4
23	0.43	0.39	0.43	0.39	0.8	0	WC	1	0.4
24	0.1	0.1	0.1	0.09	0	0	TM	2	0
25	0.11	0.1	0.09	0.08	0	0.2	WC	2	0.4
26	0.16	0.15	0.1	0.09	0.2	0	WC	2	0
27	0.11	0.13	0.1	0.13	0.2	0	WC	2	0
28	0.1	0.13	0.08	0.07	0	0	WC	2	0
29	0.07	0.06	0.07	0.06	0	0	WC	2	0.2
30	0.08	0.07	0.06	0.05	0	0	WC	2	0.2
31	0.08	0.07	0.08	0.03	0.4	0.4	TM	3	0.4

Continued on next page

Table B.3 – continued from previous page

No	System Readings				Survey Readings		Rel	HC	E_M
	Strongest Path Algo		Shortest Path Algo						
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
32	0.02	0.09	0.02	0.05	0	0.2	WC	3	0
33	0.22	0.18	0.12	0.03	0.2	0.4	WC	3	0.4
34	0.05	0.05	0.04	0.04	0	0.2	WC	3	0.6

Table B.4: Trust ratings from system and survey with each row representing meansurements between a pair of participants, $trust^{eprints}$ represents proxy trust data from ePrints co-authorship network and $trust^{mudi}$ are the trust metrics from consolidated version of ePrints and WAIS. Rel and E_M shows relationship and expertise match between participants, PL represents path length between participants in co-authorship network. SP in the Rel represents Supervisor, TM is the abbreviation of Team Member and WC shows those WAIS research group colleagues.

No	System Readings				Survey Readings		Rel	PL	E_M
	Strongest Path Algo		Shortest Path Algo						
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
1	0.3	0.27	0.3	0.27	0.8	0.4	TM	1	0.2
2	0.24	0.22	0.24	0.22	0.6	0	TM	1	0.2
3	0.15	0.14	0.15	0.14	0.6	0.6	TM	1	0.2
4	0.03	0.03	0.02	0.02	0.6	0	SP	2	0
5	0.02	0.02	0.07	0.06	0.4	0.6	TM	2	0.6
6	0.06	0.04	0.04	0.03	0	0	WC	2	0.2
7	0.04	0.02	0.03	0.01	0.2	0	WC	3	0.2
8	0.01	0.01	0.01	0.007	0	0	WC	3	0.2

Continued on next page

Table B.4 – continued from previous page

No	System Readings				Survey Readings		Rel	PL	E_M
	Strongest Path Algo		Shortest Path Algo						
	$trust^{eprints}$	$trust^{mudi}$	$trust^{eprints}$	$trust^{mudi}$	$trust^{past}$	$trust^{future}$			
9	0.02	0.02	0.02	0.02	0	0	WC	3	0
10	0.02	0.01	0.01	0.009	0	0	WC	3	0
11	0.02	0.02	0.02	0.01	0	0	EC	3	0

Appendix C

UNIVERSITY OF Southampton

ERGO application form – Ethics form

All mandatory fields are marked (M*). Applications without mandatory fields completed are likely to be rejected by reviewers. Other fields are marked "if applicable". Help text is provided, where appropriate, in italics after each question.

1. APPLICANT DETAILS

1.1 (M*) Applicant name:	**Muhammad Imran**
1.2 Supervisor (if applicable):	**Dr. David Millard, Dr. Thanassis Tiropanis**
1.3 Other researchers/collaborators (if applicable): ***Name, address, email, telephone***	

2. STUDY DETAILS

2.1 (M*) Title of study:	**Study to analyse the accuracy of aggregated trust measures from consolidated multiple social networks**
2.2 (M*) Type of study (*e.g. Undergraduate, Doctorate, Masters, Staff*):	**Doctorate**
2.3 i) (M*) Proposed start date:	**15/11/2012**
2.3 ii) (M*) Proposed end date:	**15/12/2012**

2.4 (M*) What are the aims and objectives of this study?
This study aims to find the actual trust that participants of the wais projects and eprints networks hold about each other in the professional life and to analyse it in comparison with measurements generated by the simulation study held separatly.

2.5 (M*) Background to study (*a brief rationale for conducting the study*):
This study is a part of PhD research that explores the potential idea of generating aggregated trust measurements between individuals on the web by consolidating multiple distributed (MuDi) social networks. It is now very common for users on the web to become part of MuDi social networks because not only these different types of networks serve different purposes but also gives an opportunity to interact participants from different background as compared to single networks. Activities and interactions of users in these networks provide us the opportunity to asses and integrate trust data available in MuDi social networks to generate aggregated trust values and we claim that resultant values better reflect the perspective of trust that people hold about each other in the real life.

2.6 (M*) Key research question (*Specify hypothesis if applicable*):
Our hypothesis is that aggregated trust measurements from MuDi networks can

UNIVERSITY OF Southampton

help us improve existing trust applications in two ways:

•It will shift the inter-personal trust between users on the web to the level that users perceive about each other in actual unlike existing techniques to base trust on single networks.
•Trust aware expert recommendation can help generating personalised list of experts unlike global ranking of experts in existing expert finding algorithms.

2.7 (M*) Study design (*Give a brief outline of basic study design*)
Outline what approach is being used, why certain methods have been chosen.

Design of the experiment involves describing two sub components, selection of the rating/rated participants and questions asked in the survey.

Selection of Participants:

There are two aspects of participant selection in our survey, first set of people are those participating (rating participants) in the survey and the second set are selected group of people (rated participants) about which rating participants will express their implied trust. Selection of rating participants is probabilistic within eprints and iamrearcher domain as anyone in the dataset have equal opportunity to become part of the survey but they need to LogIn first (using firrstname, lastname, research interests), while selection of the rated participants uses information from the dedicated egocentric network extracted with logged-in user as an ego. As decay of trust exists along trust path in the simulated model of trust so the trust measurements for indirectly connected participants also needs to be recorded to judge the significance of this claim. Keeping this in mind, rated participants are allegdly selected belonging to different path lengths from the rating participants.

Set of Questions:

As we are going to measure the quality of consolidated trust in a pair of MuDi professional social networks, that includes co-authorship and friendship networks, this survey aims to extract the proxy trust that participants infer about each other in the professional context. Each of the participating user will be presented with a set of questions which will help us collect implied trust data about other network members and a list of ranked experts related to his field. Later on, these measurements would be comapred to those from simulation studies to evaluate the performance of our simulated trust consolidation model.

Substance of the survey and especially selection of the questions is also tricky in studies involving human participation as it is never easy to mine relevant information from the mind of the people. Assurance of the data privacy and ethics while asking these questions increases the challange if discussed in the context of trust as it aims to get one's personal sentiments about others. Proxy trust evaluation survey in our case also tries to extract such information and models it using two questions asking people about their past work experience with the person and liklehood of working in the future if there would be any such oppourtunity available. Users can select one out of the five options available accross both the questions and those selections
would be stored in the database as integer values in the range 1 to 5 (corresponds to
left to right in the mockup) with increments of 1.

Looking further into the academic and professional networks between participants, links between users can be divided into multiple categories

UNIVERSITY OF Southampton

depending upon their roles in these networks e.g. supervisor, colleague etc as this will give us oppourtunity to analyse trust in each of these categories particularly. Forth question in the survey serves that purpose and asks users to briey explain their relationship with each of the rated person separatly and later on codification of these different type of relationships would generate set of relationship categories. Analysis held for all type of relationships generically in the above hypothesis would now be repeated for people from each of these categories specifically analysing simulated measurement of which category shifts more closely to the real life evaluations. **Another potential benefit of MuDi networks consolidation is to improve existing expert recommendation mechanisms, by presenting better ranking and more options of experts as compared to individual social networks. Existing expert finding systems considers research profile along with number of directly related experts as the metric to classify global list of experts in any domain, while our hypothesis is that trust aware personalised expert ranking using consolidated MuDi networks from different domains allows better options of experts than individual networks. This is because it not only combines information about users from multiple networks but also includes more experts that are not part of individual networks. To test this hypothesis, research interests of each rating participant are taken as an input and then a list of experts from both the networks related to that field is presented, requesting user to rank these experts. Ranked list submitted by each rating participant is stored in the database and then comapred with the other three lists generated by the simulation (one from each of the individual networks and other one from consolidated MuDi networks) using rank correlation mechanism.**

3. SAMPLE AND SETTING

3.1 (M*) How are participants to be *approached*? *Give details of what you will do if recruitment is insufficient. If participants will be accessed through a third party (e.g. children accessed via a school) state if you have permission to contact them and upload any letters of agreement to your submission in ERGO*.
I am expecting atleast 50 participants in this survey but a comprehensive email would be cucirculated among potential participants to convince maximum users become part of this study. As past experience shows that people feel hesitant to participate in any survey/study and in this case it is expected specifically because it is related to getting personal consent about others, so our plan is to attract participants mentioning draw of the Amazon vouchers.

3.2 (M*) Who are the proposed sample and where are they from (e.g. fellow students, club members)? *List inclusion/exclusion criteria if applicable*. *NB The University does not condone the use of 'blanket emails' for contacting potential participants (i.e. fellow staff and/or students).* *It is usually advised to ensure groups of students/staff have given prior permission to be contacted in this way, or to use of a third party to pass on these requests. This is because there is a potential to take advantage of the access to 'group emails' and the relationship with colleagues and subordinates; we therefore generally do not support this method of approach.* *If this is the only way to access a chosen cohort, a reasonable compromise is to obtain explicit approval from the Faculty Ethics Committee (FEC) and also from a senior member of the Faculty in case of complaint.*
Proposed sample for this survey can be anyone from two professional networks iamresearcher or eprints and hence it can include supervisors, teachers,

colleagues, university fellows etc. Potential participants would be identified and a preliminary email would be circulated to them.

3.3 (M*) Describe the relationship between researcher and sample (*Describe any relationship e.g. teacher, friend, boss, clinician, etc.*)
University fellow

3.4 (M*) Describe how you will ensure that fully informed consent is being given: (*include how long participants have to decide whether to take part*)
At first instance, all the required information will be provided in the introdcutory email so that people can get to know everything about the study. Similar information would also be provided at the Log In page of the survey portal as a reminder. Furthremore, they also need to tick the checkbox to inform their consent before the start of questionnaire.

4. RESEARCH PROCEDURES, INTERVENTIONS AND MEASUREMENTS

4.1 (M*) Give a brief account of the procedure as experienced by the participant *(Make clear who does what, how many times and in what order. Make clear the role of all assistants and collaborators. Make clear total demands made on participants, including time and travel). Upload any copies of questionnaires and interview schedules to your submission in ERGO.*
The participant needs to follow the folowing steps for completing the survey: Step 1: In the introductory email all the relevant information related to study will be provided for participant to decide whether or not to participate in the study. Step2: In case of willingness, they need to follow the online survey by clicking on the link provided. Step 3: First page of the survey will ask participate to enter firstname, lastname and research area as this will generate personalised survey material having a list of 8 people to rate and a list of 8 experts to rank. This information will be different for each person and will be based on his connections in the network(s) and research area. Input information from this step will not be stored in the database. Step 4: After answering all the questions user will click on the SUBMIT button to finish the survey and the information will be stored in the MySQL database. It will take at maximum 10 minutes to complete this survey.

5. STUDY MANAGEMENT

5.1 (M*) State any potential for psychological or physical discomfort and/or distress?
There is no psychological or physical discomfort associated with this study.

UNIVERSITY OF Southampton

5.2 (M*) Explain how you intend to alleviate any psychological or physical discomfort and/or distress that may arise? (if applicable)
N/A

5.3 Explain how you will care for any participants in 'special groups' *(i.e. those in a dependent relationship, vulnerable or lacking in mental capacity)* (if applicable)?
N/A

5.4 Please give details of any payments or incentives being used to recruit participants (if applicable)?
We are planning to conduct draw of Amazon vouchers as an incentive.

5.5 i) How will participant anonymity and/or data anonymity be maintained (if applicable)? ***Two definitions of anonymity exist:*** ***i)* Unlinked anonymity - *Complete anonymity can only be promised if questionnaires or other requests for information are not targeted to, or received from, individuals using their name or address or any other identifiable characteristics. For example if questionnaires are sent out with no possible identifiers when returned, or if they are picked up by respondents in a public place, then anonymity can be claimed. Research methods using interviews cannot usually claim anonymity – unless using telephone interviews when participants dial in.*** ***ii)* Linked anonymity *- Using this method, complete anonymity cannot be promised because participants can be identified; their data may be coded so that participants are not identified by researchers, but the information provided to participants should indicate that they could be linked to their data.***
Linked anonymity technique would be used to hide identity of the participants, so that even researcher cannot identify what has been said by someone. However information shared by participants would help us create links between anonymised entities.

5.5 ii) How will participant confidentiality be maintained (if applicable)? ***Confidentiality is defined as the non-disclosure of research information except to another authorised person. Confidential information can be shared with those who are already party to it, and may also be disclosed where the person providing the information provides explicit consent.***
Data collected from this survey will be confidential accessible only to researcher and noway would be disclosed to any other person.

5.6 (M*) How will personal data and study results be stored securely during and after the study? *Researchers should be aware of, and compliant with, the Data Protection policy of the University. You must be able to demonstrate this in respect of handling, storage and retention of data.*
Data would be stored in MySQL database using linked anonymisation technique and randomly generated code would be replaced by IDs to hide identity of the participants.

5.7 (M*) Who will have access to these data?
Only researcher can access the collected data for analysis purpose and later on it would be disposed off as well.

UNIVERSITY OF Southampton

N.B. – Before you upload this document to your ERGO submission remember to:

1. Complete ALL mandatory sections in this form

2. Upload any letters of agreement referred to in question 3.1 to your ERGO submission

3. Upload any interview schedules and copies of questionnaires referred to in question 4.1

Figure C.1: Ethics Form

UNIVERSITY OF Southampton

Project Description (Protocol)

Study Title: Study to analyse the accuracy of aggregated trust measures from consolidated multiple social networks

Researcher(s) Muhammad Imran

Background

This PhD research explores the potential idea of generating aggregated trust measurements between individuals on the web by consolidating multiple distributed (MuDi) social networks. It is now very common for users on the web to become part of MuDi social networks because not only these different types of networks serve different purposes but also gives an opportunity to interact participants from different background as compared to single networks. Activities and interactions of users in these networks provide us opportunity to asses and integrate trust data available in MuDi social networks to generate aggregated trust values and our hypothesis is that resulted measurements can help us improve existing algorithms in two ways:

- It will shift the inter-personal trust between users on the web to the level that users perceive about each other in actual unlike existing techniques to base trust on single social networks.
- Trust aware expert recommendation can help generating personalised list of experts as compared to global ranking of experts in existing expert finding algorithms.

Method

We are planning to conduct an online survey to collect real life quantitative and qualitative data about inter-personal professional trust and expert recommendation scenario. RDF/XML user data extracted from two professional social networks, eprints co-authorship network extracted from eprints publication network and wais collaboration network is consolidated and participating users from either or both the networks need personal attributes (firstname, lastname, research interests) to LogIn to the survey portal. Each of the logged in user will be redirected to a single page personalised survey generated using MuDi RDF graph asking set of questions about related people and ranking a list of experts related to his research area. After successful completion user will click the submit button to finish the survey and all the data will be saved in the MySQL database.

Material

Material of the survey is divided in two parts each dedicated to one of the hypothesis point, first part asks list of 4 questions about each of the 8 selected users and second part includes ranking a list of 8 experts.

In both parts of the survey, if the user exists in both the networks then related people are randomly selected with proportion of 4:4 from eprints and wais projects networks respectively with variable path length of 1, 2 and 3 from rating participant. But if the user only exists in one of the network then all 8 users are selected from that respective network.

[Date: 21/10/12] [Version number: 1]

UNIVERSITY OF Southampton

First two questions focus on getting trust information between users asking participant about their past experience with the person if he ever worked with him in the past and his intention to work with that person if he ever gets a chance in future. Third question asks about alignment between expertise of rating and rated participant as that can help us understand correlation between their professional relationship and expertise. Forth question asks participant to briefly explain their relationship with each of the rated person separately, e.g. supervisor, colleague etc, later on codification of these different types of relationships would let us generate set of relationship categories. Analysis held for all types of relationships generically using questions 1 and 2 would now be repeated for people from each of these categories specifically analysing simulated measurement of which category shifts more closely to the real life evaluations. The last question is specifically about expert recommendation and requests participant to rank list of experts related to his field keeping in view his preference to interact someone related to his research area.

All the questions including multiple choice questions about trust are quite flexible, easy to understand and guides users at each step of answering without creating any confusion. For example, it asks participant about his familiarity with the person before asking any specific question related to trust. Furthermore, efforts have been made to use simple English and to keep the maximum completion time within 10 minutes.

Participants

Anyone from those parts of wais projects or eprints networks can participate in the survey but set of participants would be selected first and the link of introductory email would only be circulated among them. As past experience shows that people feel hesitant to be part of any study and in this case it is expected specifically because it is related to getting personal consent about others, so our plan is to attract participants mentioning draw of the Amazon vouchers.

Procedure

Soon after approval of the survey material by the ethics committee, there is a plan to upload the survey on a university web server 24/7 for a period of one month. Users will fill in that survey by following the link mentioned in the circulated email and data will be stored in MySQL database at the back end. In the meantime there would be no involvement from the researcher side and users will be free to reply the way they like.

Statistical analysis

There would be two set of analysis on the data generated by this survey, first one is to assess the accuracy of the aggregated trust produced by the consolidated version of MuDi social networks and second is to measure the accuracy of trust aware expert recommendation.

For each of the analysis phase, there would be in total four set of results after conducting simulation (help separately) and survey studies; three from numerical study constituting data corresponding to two individual networks and a consolidated network and a set from this survey. Comparative analysis among the corresponding datasets will be performed to judge the level of similarity between simulated and real world data. This would help us in making any final decision whether consolidation of MuDi networks is anyway better approach for trust and expert recommendation.

[Date: 21/10/12] [Version number: 1]

Figure C.1: Project Description

UNIVERSITY OF Southampton

FPAS Participant Information Sheet

Study Title: Study to analyse the accuracy of aggregated trust measures from consolidated multiple social networks

Researcher: Muhammad Imran, ECS, University of Southampton

Ethics Reference Number:

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

This PhD research explores the potential idea of generating aggregated trust measurements between individuals on the web by consolidating multiple distributed (MuDi) social networks. It is now very common among users on the web to become part of MuDi social networks because not only these different types of networks serve different purposes but it also gives an opportunity to interact participants from different background as compared to single networks. Activities and interactions of users in these networks provide us opportunity to combine and asses trust data available from MuDi social networks to generate aggregated trust values and we claim that these evaluations better matches the trust that users perceive in actual.

Why have I been chosen?

As this study includes people from eprints and wais projects networks and you being part of any or both of these networks are humbly requested to be part of it.

What will happen to me if I take part?

This exercise is just an online questionnaire aiming to get proxy trust information between individuals in professional social networks and to rank a list of experts related to your research interests. It will take at maximum 10 minutes and involves answering set of questions about related people and ranking a set of experts related to your field.

Are there any benefits in my taking part?

Yes, there would be a draw of Amazon Vouchers and you have got the equal chance of being among the winners. Greater service would be for the web science community to make web more robust and secure place for its users.

Are there any risks involved?

As this study is approved by the Ethics committee and respects all the standards of Data Protection Act, so information would be totally confidential, accessible only to researchers, hence will not result in any unforeseen problems for you.

Will my participation be confidential?

UNIVERSITY OF Southampton

Yes, security of the information is ensured by storing it on a password protected computer and using linked anonymity technique where identity of the person is replaced with a certain random code which researchers themselves cannot track even. Furthermore, this information will be confidential and only researcher has the right to use that information and after use it would be destroyed as well.

What happens if I change my mind?

If you feel like no more interested to complete this survey then you can withdraw anytime by closing the browser unless you have submitted by clicking SUBMIT button. Also this would not result saving your data on our system as it is saved after you finish the survey. In case you are interested again; you need to fill in again before submission.

What happens if something goes wrong?

In case of any problem, you can contact the chair of Ethics Committee, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Where can I get more information?

In case of any question or query, please free to contact me, (mi1g08@ecs.soton.ac.uk).

[Oct 2012] [Version number: 1]

Figure C.1: Participant Information Sheet

UNIVERSITY OF Southampton

CONSENT FORM (*Insert Version number*)

Study title: Study to analyse the accuracy of aggregated trust from multiple social networks

Researcher name: Muhammad Imran
Ethics reference number:

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (insert date /version no. of participant information sheet) and have had the opportunity to ask questions about the study.	☐
I agree to take part in this research project and agree for my data to be used for the purpose of this study	☐
I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected	☐
I am happy to be contacted regarding other unspecified research projects. I therefore consent to the University retaining my personal details on a database, kept separately from the research data detailed above. The 'validity' of my consent is conditional upon the University complying with the Data Protection Act and I understand that I can request my details be removed from this database at any time.	☐

Data Protection
I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name)……………………………………………………

Signature of participant……………………………………………………………..

Date…………………………………………………………………………………

[24-10-2012] [1]

UNIVERSITY OF Southampton

Tips for Designing a Consent Form

(Please delete this section before submitting your documents to your Ethics Committee/RGO)

Participant consent must be received in writing using a signed consent form, which may be electronic. Participants must be able to (print off and) take away (or store) a copy of their signed consent form.

The main features of a good consent form are:

Date and Version number
It is important that the consent form is version numbered and dated so it is possible to track changes if and when they occur.

Use of Ethics reference
This is evidence of ethical approval and will reassure participants – enter the Submission ID generated when you create a submission in ERGO

Use of itemised statement to allow each component of the research to be agreed to
Information will be commensurate with the study. For example, in an interview study you may want consent to (i) interview and (ii) tape the interview.

Use of Initial boxes
In general, participants should initial, NOT tick, any consent form boxes, to minimise fraud. Do not provide 'pre-ticked' consent forms. In an on-line form, boxes may be ticked if it is clear that only the participant can tick them.

Use of participant in other research
If you wish to keep the contact details of the participant for potential use in further studies you should include a separate statement for them to initial to give consent, and be clear that they can be removed from this contact list at any time.

Confirmation of the right to withdraw
You may wish to include a separate statement on confidentiality/anonymity but this is often best explained in the participant information sheet

Space for printed names, signatures and dates
A space for the name and signature of the person taking consent is also desirable if different from the named researcher

For studies involving the NHS
For NHS research, extensive guidance notes and exemplars are available on the National Research Ethics Support website:
http://www.nres.npsa.nhs.uk

For studies involving minors/vulnerable adults
For studies involving minors/vulnerable adults, consent should be taken from the parent/guardian/carer and it is desirable for the participant to sign an assent form to indicate their willingness to take part. There are situations where it is appropriate to use 'opt-out consent' (informing parents/carers of the study and that if they do not respond to inform the researcher that they *do not* want their child/dependent to take part then it will be assumed that their

UNIVERSITY OF Southampton

consent is given). If you are in any doubt about the method of consent required you should seek advice from your local Ethics Committee or the Research Governance Office.

[24-10-2012] [1]

Figure C.1: Consent Form

Appendix D

This focus group study was conducted to evaluate first draft of the trust questionnaire. It asked following set of questions to a group of 10 people about the number of questions they can easily answer if they are presented with such questionnaire, best way of presenting these questions (online/ paper-form), ethical issues that can arise in the meantime and so on.

FOCUS GROUP SESSION

Please answer the following questions by analysing set of questions mentioned in the survey.

Is it easy to answer the questions with current layout of the survey (please imagine as if all the questions are presented on a single web page)?

Comments

How sensitive are the questions with respect to ethic standards.

Comments

How likely it is that rating participant can correctly rate if merely name of the rated person is mentioned.

Comments

The presented survey is about measuring implied trust that people hold about each other in professional social network. Now please answer following questions by analyzing the survey once again.

Please answer the following questions by analysing set of questions mentioned in the survey.

Do you feel that the questions asked in the survey adequately represent or percept person to person trust scenario in professional context.

Comments

What is your consent about number of questions and which of the other potential questions can be included.

Comments

How long it takes to complete the survey and what is the other convenient way of presenting these questions.

Comments

Do these questions allow participants to express their sentiments without any threat to personal relationships?

Comments

What would be the best incentive for attracting people to become part of the survey?

Comments

How likely it is that rating participant can correctly rate if merely name of the rated person is mentioned.

Comments

Figure D.0: Questionnaire presented to a focus group session

Bibliography

Ahn, Y.-Y., Han, S., Haewoon, K., Moon, S., and Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In *Proceedings of 16th international conference on World Wide Web(WWW)*, pages 835–844, Banff, Alberta, Canada. ACM.

Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152.

Avesani, P., Massa, P., and Tiella, R. (2005). Moleskiing. it: a trust-aware recommender system for ski mountaineering. *International Journal for Infonomics*, 20(35):1–10.

Bae, J. and Kim, S. (2009). A global social graph as a hybrid hypergraph. In *5th Joint International Conference on Network Computing (INC), International Conference on Advanced International Management and Services (IMS) and International Conference on Digital Content, Multimedia Technology and its Applications (IDC)*, pages 1025–1031, Seoul, South Korea. IEEE Computer Society.

Barrat, A. and Weigt, M. (2000). On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, 13:547–560.

Berners-Lee, T. (1998). The semantic web roadmap. http://www.w3.org/DesignIssues/Semantic.html. [Last accessed on 21-June-2014].

Berners-Lee, T. (2006). Linked data. http://www.w3.org/DesignIssues/LinkedData.html. [Last accessed on 21-June-2014].

Berners-Lee, T., Hendler, J., and Lassila (2001). The semantic web. *Scientific American*, 284:34–43. http://www.scientificamerican.com/article/the-semantic-web/.

Bilge, L., Strufe, T., Balzarotti, D., and Kirda, E. (2009). All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of 18th International Conference on World Wide Web (WWW)*, pages 551–560, Madrid, Spain. ACM.

Bistarelli, S. and Santini, F. (2014). Two trust networks in one: Using bipolar structures to fuse trust and distrust. In *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*, pages 383–390.

Bizer, C. and Cyganiak, R. (2013). TriG. http://www.w3.org/TR/2013/WD-trig-20130409/. [Last accessed on 21-June-2014].

Board, D. U. (2012). Dublin core metadata initiative terms. http://dublincore.org/documents/dcmi-terms/. [Last accessed on 21-June-2014].

Boyd, D. and Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230.

Brickley, D. and Guha, R. V. (2004). Resource description framework schema (RDFS) - W3C recommendation. http://www.w3.org/TR/rdf-schema/. [Last accessed on 21-June-2014].

Brickley, D. and Miller, L. (2010). FOAF vocabulary specification namespace. http://xmlns.com/foaf/spec/. [Last accessed on 21-June-2014].

Broekstra, J., Kampman, A., and Harmelen, F. (2002). Sesame: A generic architecture for storing and querying RDF and RDF schema. In *1st International Semantic Web Conference (ISWC)*, volume 2342 of *Lecture Notes in Computer Science*, pages 54–68, Sardinia, Italy. Berlin/Heidelberg: Springer.

Broekstra, J., Kampman, A., and Van Harmelen, F. (2003). Sesame: An architecture for storing and querying RDF data and schema information. *Semantics for the World Wide Web (WWW)*, page 197.

Carroll, J., Bizer, C., Hayes, P., and Stickler, P. (2005a). Named graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(4):247–267.

Carroll, J., Bizer, C., Hayes, P., and Stickler, P. (2005b). Named graphs, provenance and trust. In *Proceedings of 14th International Conference on World Wide Web (WWW)*, pages 613–622, Chiba, Japan. ACM.

Carroll, J. J. and Stickler, P. (2004). RDF triples in XML. In *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers and Posters*, pages 412–413, New York, USA.

Chakraborty, P. S. and Karform, S. (2012). Designing trust propagation algorithms based on simple multiplicative strategy for social networks. *Procedia Technology*, 6:534 – 539. 2nd International Conference on Communication, Computing and Security [ICCCS-2012].

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94:95–120.

Cook, J. and Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfilment. *Journal of Occupational Psychology*, 53(1):39–52.

Corritore, C., Kracher, B., and Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6):737–758.

Crowder, R., Hughes, G., and Hall, W. (2002a). An agent based approach to finding expertise. In *4th International Conference on Practical Aspects of Knowledge Management*, pages 179–188, Vienna, Austria. Berlin/Heidelberg: Springer.

Crowder, R., Hughes, G., and Hall, W. (2002b). Approaches to locating expertise using corporate knowledge. *Intelligent Systems in Accounting, Finance and Management*, 11(4):185–200.

Cyganiak, R., Harth, A., and Hogan, A. (2012). N-quads: Extending n-triples with context. http://sw.deri.org/2008/07/n-quads/. [Last accessed on 21-June-2014].

De Bruijn, J., Ehrig, M., Feier, C., Martin-Recuerda, F., Scharffe, F., and Weiten, M. (2006). Ontology mediation, merging and aligning. In *Semantic Web Technologies. Trends and Research in Ontology-based Systems*, Chichester, UK. Wiley and Sons.

Dodds, L. and Davis, I. (2011). Linked data patterns - A pattern catalogue for modelling, publishing, and consuming linked data. http://patterns.dataincubator.org/book/. [Last accessed on 21-June-2014].

Doerr, B., Fouz, M., and Friedrich, T. (2012). Why rumors spread so quickly in social networks. *Commun. ACM*, 55(6):70–75.

Dou, D., McDermott, D., and Qi, P. (2002). Ontology translation by ontology merging and automated reasoning. In *Proceedings of European Knowledge Acquisition (EKAW) Workshop on Ontologies for Multi-Agent Systems*, pages 73–94, Siguenza, Spain. Berlin/Heidelberg: Springer.

ECS (2013). ECS ontology. http://rdf.ecs.soton.ac.uk/ontology/ecs. [Last accessed on 12-September-2013].

Filev, D. and Yager, R. (1994). Learning OWA operator weights from data. In *Proceedings of 3rd IEEE Conference on Fuzzy Systems, World Congress on Computational Intelligence*, pages 468–473, Orlando, FL, USA.

Fukuyama (1995). *Trust: The Social Virtues and the Creation of Prosperity*. The Free Press.

Futrelle, J. (2006). Harvesting rdf triples. In *Provenance and Annotation of Data*, volume 4145 of *Lecture Notes in Computer Science*, pages 64–72. Berlin/Heidelberg: Springer.

Garton, L., Haythornthwaite, C., and Wellman, B. (1997). Studying online social networks. *Journal of Computer-Mediated Communication*, 3(1).

Glaser, H., Jafri, A., and Millard, I. (2009). Managing co-reference on the semantic web. In *Linked Data on the Web (LDOW) Workshop with 18th International Conference on World Wide Web (WWW)*, Madrid, Spain.

Glaser, H., Millard, I., Jaffri, A., Lewy, T., Millard, I., and Dowling, B. (2008). On coreference and the semantic web. In *Proceedings of 7th International Semantic Web Conference (ISWC)*, pages 26–30, Karlsruhe, Germany.

Golbeck, J. (2005). *Computing and Applying Trust in Web-Based Social Networks*. PhD thesis, Department of Mathematics and Computer Science, University of Maryland, US.

Golbeck, J. (2006). Trust on the world wide web: A survey. *Foundations and Trends in Web Science.*, 1(2):131–197.

Golbeck, J. and Hendler, J. (2006a). Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of 3rd IEEE Consumer Communications and Networking Conference (CCNC)*, volume 1, pages 282–286, Las Vegas, NV, USA.

Golbeck, J. and Hendler, J. (2006b). Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology.*, 6(4):497–529.

Golbeck, J. and Parsia, B. (2006). Trust network-based filtering of aggregated claims. *International Journal of Metadata, Semantics and Ontologies*, 1(1):58–65.

Golbeck, J., Parsia, B., and Hendler, J. (2003). Trust networks on the semantic web. In *Proceedings of 7th International Workshop, Cooperative Information Agents (CIA) VII*, volume 2782 of *Lecture Notes in Computer Science*, pages 238–249. Berlin/Heidelberg: Springer, Helsinki, Finland.

Grandison, T. and Sloman, M. (2000). A survey of trust in internet applications. *IEEE Communications Surveys Tutorials*, 3(4):2–16.

Griffiths, N., Chao, K.-M., and Younas, M. (2006). Fuzzy trust for peer-to-peer systems. In *Proceedings of 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCS)*, page 73, Lisboa, Portugal.

Guha, R. (2003). Open rating systems. In *Technical report, Stanford Knowledge Systems Laboratory*, Stanford, CA, USA. http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/open_rating_systems/wot.pdf.

Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In *Proceedings of 13th International Conference on World Wide Web (WWW)*, pages 403–412, Rio De Janeiro, Brazil. ACM.

Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., and Farrell, S. (2008). Harvesting with SONAR: The value of aggregating social network information. In *Proceeding of 26th Annual Special Interest for Computer-Human Interaction (SIGCHI)*

Conference on Human Factors in Computing Systems (CHI), pages 1017–1026, Florence, Italy. ACM.

Hayes, P. (2004). RDF semantics. http://www.w3.org/TR/rdf-mt/#ReifAndCont. [Last accessed on 21-June-2014].

Heath, T. and Motta, E. (2008). The Hoonoh ontology for describing trust relationships in information seeking. In *Proceedings of 3rd Personal Identification and Collaborations: Knowledge Mediation and Extraction (PICKME) workshop with 7th International Conference on Web Semantics*, pages 67–75, Karlsruhe, Germany. Berlin/Heidelberg: Springer.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, USA.

Hogan, A., Harth, A., and Decker, S. (2007). Performing object consolidation on the semantic web data graph. In *Proceedings of I3; Identity, Identifiers, Identification, in conjunction with 16th International World Wide Web Conference (WWW)*, Banff, Alberta, Canada. ACM.

Holland, P. and Leinhardt, S. (1972). Some evidence on the transitivity of positive interpersonal sentiment. *American Journal of Sociology*, 77(6):1205–1209.

Hwang, S.-Y., Wei, C.-P., and Liao, Y.-F. (2010). Coauthorship networks and academic literature recommendation. *Electronic Commerce Research and Applications*, 9(4):323–334.

Imran, M., Millard, D., and Tiropanis, T. (2012). Impact of consolidating social networks on derived trust factors. *ASE Human Journal*, 1(2):88–99.

Josang, A., Gray, E., and Kinateder., M. (2003). Analysing topologies of transitive trust. In *Proceedings of the Workshop of Formal Aspects of Security and Trust (FAST)*, Pisa, Italy. ftp://ftp.informatik.uni-stuttgart.de/pub/library/ncstrl.ustuttgart_fi/INPROC-2003-19/INPROC-2003-19.pdf.

Jung, J. and Euzenat, J. (2007). Towards semantic social networks. In *Proceedings of 4th European Semantic Web Conference (ESWC)*, volume 4519 of *Lecture Notes in Computer Science*, pages 267–280, Innsbruck, Austria. Berlin/Heidelberg: Springer.

Kalfoglou, Y. and Schorlemmer, M. (2003). Ontology mapping: The state of the art. *The Knowledge Engineering Review*, 18(1):1–31.

Kamvar, S., Schlosser, M., and Garcia-Molina, H. (2003). The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of 12th International Conference on World Wide Web (WWW)*, pages 640–651, Budapest, Hungary. ACM.

Kim, Y. A. and Song, H. S. (2011). Strategies for predicting local trust based on trust propagation in social networks. *Knowledge-Based Systems*, 24(8):1360 – 1371.

Kleyn, G. and Carroll, J. J. (2004). Resource description framework (rdf): Concepts and abstract syntax - W3C recommendation. http://www.w3.org/TR/rdf-concepts/. [Last accessed on 21-June-2014].

Koivunen, M.-R. and Miller, E. (2001). W3C semantic web activity. page 2744. http://www.w3.org/2001/12/semweb-fin/w3csw.

Lawrence, P., Sergey, B., Rajeev, M., and Terry, W. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Leinhardt, S. (1972). Developmental change in the sentiment structure of children's groups. *American Sociological Review*, 37(2):202–212.

Lesani, M. and Bagheri, S. (2006). Applying and inferring fuzzy trust in semantic web social networks. In *Canadian Semantic Web*, volume 2 of *Semantic Web and Beyond*, pages 23–43. Berlin/Heidelberg: Springer.

Levien, R. (2009). Attack-resistant trust metrics. In *Computing with Social Trust*, Human-Computer Interaction, pages 121–132. London: Springer.

Lewis, D. and Weigert, A. (1985). Trust as a social reality. *Social Forces*, 63(4):967–985.

Li, J., Tang, J., Zhang, J., Luo, Q., Liu, Y., and Hong, M. (2007). EOS: Expertise oriented search using social networks. In *Proceedings of 16th International Conference on World Wide Web (WWW)*, pages 1271–1272, Banff, Alberta, Canada. ACM.

Liu, G., Wang, Y., and Orgun, M. A. (2010). Optimal social tust path selection in complex social networks. *Proceedings of 24th Autonomous Agents and Artificial Intelligence (AAAI) Conference on Artificial Intelligence*, pages 1391–1398. http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/viewFile/1751/2218.

Liu, J., Liu, D., Yan, X., Dong, L., Zeng, T., Zhang, Y., and Tang, J. (2014). Aminer-mini: A people search engine for university. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM 14, pages 2069–2071, New York, NY, USA. ACM.

Ma, Y., Lu, H., Gan, Z., and Ma, X. (2014). Trust discounting and trust fusion in online social networks. In *Web Technologies and Applications*, volume 8709 of *Lecture Notes in Computer Science*, pages 619–626. Springer International Publishing.

Matsuo, Y., Hamasaki, M., Nakamura, Y., Nishimura, T., Hasida, K., Takeda, H., Mori, J., Bollegala, D., and Ishizuka, M. (2006). Spinning multiple social networks for semantic web. In *Proceedings of the 21st National Conference on Artificial Intelligence*, volume 2, pages 1381–1386, Boston, Massachusetts. AAAI Press.

McGuinness, D. L. and van Harmelen, F. (2004). Ontology web language (OWL) - W3C recommendation. http://www.w3.org/TR/owl-features/. [Last accessed on 21-June-2014].

Mika, P. (2005). Flink: Semantic web technology for the extraction and analysis of social networks. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2-3):211–223.

Mislove, A., Marcon, M., Gummadi, K., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM Special Interest Group on Data Communication (SIGCOMM) Conference on Internet Measurement(IMC)*, pages 29–42, San Diego, California, USA.

Moyano, F., Fernandez-Gago, C., and Lopez, J. (2012). A conceptual framework for trust models. In *Trust, Privacy and Security in Digital Business*, volume 7449 of *Lecture Notes in Computer Science*, pages 93–104. Springer Berlin Heidelberg.

Mtibaa, A., May, M., Diot, C., and Ammar, M. (2010). Peoplerank: Social opportunistic forwarding. In *Proceedings of IEEE International Conference on Computer Communications(INFOCOM)*, pages 1–5, San Diego, CA.

Mui, L., Mohtashemi, M., and Halberstadt, A. (2002). A computational model of trust and reputation. In *Proceedings of 35th Annual Hawaii International Conference on System Sciences (HICSS)*, pages 188–196, Big Island, HI, USA. IEEE Computer Society.

Nepal, S., Paris, C., Bista, S. K., and Sherchan, W. (2013). A trust model-based analysis of social networks. *International Journal of Trust Management in Computing and Communications 19*, 1(1):3–22.

Nepal, S., Sherchan, W., and Paris, C. (2011). Strust: A trust model for social networks. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 841–846.

Newman, M. (2001a). Scientific collaboration networks, part I. network construction and fundamental results. *Physical. Review.*, 64(1).

Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.

Nikolov, A., Uren, V., Motta, E., and Roeck, A. (2008). Integration of semantically annotated data by the knofuss architecture. In *16th International Conference for Knowledge Engineering: Practice and Patterns (EKAW)*, volume 5268 of *Lecture Notes in Computer Science*, pages 265–274. Berlin/Heidelberg: Springer.

Noy, N. and Musen, M. (2001). Anchor-prompt: Using non-local context for semantic matching. In *Proceedings of workshop on ontologies and information sharing at the international joint conference on artificial intelligence (IJCAI)*, pages 63–70, Seattle, USA.

Olmedilla, D., Rana, O., Matthew, B., and Nejdl, W. (2006). Security and trust issues in semantic grids. In *Proceedings of the Dagsthul Seminar, Semantic Grid: The Convergence of Technologies*, Dagstuhl, Germany. http://drops.dagstuhl.de/opus/volltexte/2006/408/.

Rahm, E. and Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *International Journal on Very Large Data Bases (VLDB)*, 10(4):334–350.

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2):79–101.

Rotter, J. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26(5):443–452. http://psycnet.apa.org/psycinfo/1972-02865-001.

Rousseau, D., Sitkin, S., Burt, R., and Camerer, C. (1998). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, 23(3):393–404.

Shadbolt, N., Hall, W., and Berners-Lee, T. (2006). The semantic web revisited. *Intelligent Systems, IEEE*, 21(3):96–101.

Shariff, S. and Zhang, X. (2014). A survey on deceptions in online social networks. In *Computer and Information Sciences (ICCOINS), 2014 International Conference on*, pages 1–6.

Sherchan, W., Nepal, S., and Paris, C. (2013). A survey of trust in social networks. *ACM Comput. Surv.*, 45(4):47:1–47:33.

Shi, L., Berrueta, D., Fernandez, S., Polo, L., and Fernandez, S. (2008). Smushing rdf instances: Are alice and bob the same open source developer. In *Proceedings of 3rd Personal Identification and Collaborations: Knowledge Mediation and Extraction (PICKME) Workshop with 7th International Conference on Web Semantics*, Karlsruhe, Germany. Berlin/Heidelberg: Springer. http://data.semanticweb.org/workshop/pickme/2008/paper/main/6/html.

Sim, Y.-W. and Crowder, R. (2004). Evaluation of an approach to expertise finding. In *Practical Aspects of Knowledge Management*, volume 3336 of *Lecture Notes in Computer Science*, pages 141–152. Berlin/Heidelberg: Springer.

Sleeman, J. (2012). Online unsupervised coreference resolution for semi-structured heterogeneous data. In *11th International Semantic Web Conference*, volume 7650 of *Lecture Notes in Computer Science*, pages 457–460, Boston, MA, USA,. Berlin/Heidelberg: Springer.

Sleeman, J. and Finin, T. (2010a). Learning co-reference relations for FOAF instances. In *Proceedings of the Poster and Demonstration Session at 9th International Semantic Web Conference (ISWC)*. http://ebiquity.umbc.edu/paper/html/id/503/Learning-Co-reference-Relations-for-FOAF-Instances.

Sleeman, J. and Finin, T. (2010b). A machine learning approach to linking FOAF instances. In *Proceedings of American Association for Artificial Intelligence (AAAI) Spring Symposium on Linked Data Meets Artifical Intelligence*. AAAI Press. http://ebiquity.umbc.edu/paper/html/id/471/A-Machine-Learning-Approach-to-Linking-FOAF-Instances.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: Extraction and mining of academic social networks. In *Proceedings of 14th ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 990–998, Las Vegas, Nevada, USA.

Thirunarayan, K. and Anantharam, P. (2011). Trust networks: Interpersonal, sensor, and social. In *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pages 13–21, Philadelphia, PA, USA. IEEE.

Verbiest, N., Cornelis, C., Victor, P., and Herrera-Viedma, E. (2012). Trust and distrust aggregation enhanced with path length incorporation. *Fuzzy Sets and Systems*, 202(0):61 – 74. Theme: Aggregation Functions.

Vicenc, T. (1997). The weighted OWA operator. *International Journal of Intelligent Systems*, 12(2):153–166.

Vicenc, T. and Yasuo, N. (2007a). *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin/Heidelberg: Springer.

Vicenc, T. and Yasuo, N. (2007b). A view of averaging aggregation operators. *IEEE Transactions on Fuzzy Systems*, 15(6):1063–1067.

Victor, P., Cornelis, C., Cock, M. D., and Herrera-Viedma, E. (2011). Practical aggregation operators for gradual trust and distrust. *Fuzzy Sets and Systems*, 184(1):126 – 147. Preference Modelling and Decision Analysis (Selected Papers from {EUROFUSE} 2009).

Viljanen, L. (2005). Towards an ontology of trust. In *2nd International TrustBus Conference*, volume 3592, pages 175–184, Copenhagen, Denmark. Berlin/Heidelberg: Springer.

Walter, F., Battiston, S., and Schweitzer, F. (2008). A model of a trust-based recommendation system on a social network. *Autonomous Agents and Multi-Agent Systems*, 16(1):57–74.

Wang, Y. and Vassileva, J. (2003). Trust and reputation model in peer-to-peer networks. In *Proceedings of 3rd International Conference on Peer-to-Peer Computing (P2P)*, pages 150–157, Linkoping, Sweden. IEEE.

Wang, Y. and Vassileva, J. (2005). Bayesian network-based trust model in peer-to-peer networks. In *2nd International Workshop on Agents and Peer-to-Peer Computing (AP2PC)*, pages 23–34, Melbourne, Australia. Berlin/Heidelberg: Springer.

Watkins, E. and Nicole, D. (2006). Named graphs as a mechanism for reasoning about provenance. In *8th Asia-Pacific Web Conference*, volume 3841, pages 943–948, Harbin, China. Berlin/Heidelberg: Springer.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442. http://www.nature.com/nature/journal/v393/n6684/abs/393440a0.html.

Weaver, J. and Tarjan, P. (2013). Facebook linked data via the graph API. *Semantic Web*, 4(3):245–250. http://iospress.metapress.com/content/t2745678826v6422/.

Xiong, L. and Liu, L. (2004). Peertrust: supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):843–857.

Xu, J., Lu, Q., Li, M., and Li, W. (2015). Web person disambiguation using hierarchical co-reference model. In *Computational Linguistics and Intelligent Text Processing*, volume 9041 of *Lecture Notes in Computer Science*, pages 279–291. Springer International Publishing.

Xu, Z. S. and Da, Q. L. (2003). An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18(9):953–969.

Yager, R. and Filev, D. (1998). Operations for granular computing: mixing words and numbers. In *Proceedings of IEEE International Conference on Fuzzy Systems, IEEE World Congress on Computational Intelligence*, volume 1, pages 123–128, Anchorage, AK, USA.

Yager, R. and Filev, D. (1999). Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(2):141–150.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190.

Yager, R. R. (2003). Induced aggregation operators. *Fuzzy Sets and Systems*, 137(1):59–69.

Yager, R. R. (2004). A framework for multi-source data fusion. *Information Sciences*, 163(13):175–200.

Yager, R. R. and Kacprzyk, J. (1997). *The Ordered Weighted Averaging Operators: Theory and Applications*. Boston: Kluwer Academic.

Zhang, J., Tang, J., and Li, J. (2007). Expert finding in a social network. In *Proceedings of 12th International Conference on Database Systems for Advanced Applications(DASFAA)*, volume 4443, pages 1066–1069, Bangkok, Thailand. Berlin/Heidelberg: Springer.

Zhang, J., Tang, J., Liu, L., and Li, J. (2008). A mixture model for expert finding. In *Proceedings of the 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining(PAKDD)*, pages 466–478, Osaka, Japan. Berlin/Heidelberg: Springer.

Zhang, Y., Chen, H., Wu, Z., and Zheng, X. (2006). A reputation-chain trust model for the semantic web. In *20th International Conference on Advanced Information Networking and Applications(AINA).*, volume 2, page 5, Vienna, Austria. IEEE.

Zhang, Y. and Yu, T. (2012). Mining trust relationships from online social networks. *Journal of Computer Science and Technology*, 27(3):492–505.

Ziegler, C.-N. and Lausen, G. (2004). Spreading activation models for trust propagation. In *IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE)*, pages 83–97, Taipei, Taiwan.

Ziegler, C.-N. and Lausen, G. (2005). Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, 7(4):337–358.