Social Networks Profile Mapping using Games

Mohamed Shehab, Moo Nam Ko, Hakim Touati
Department of Software and Information Systems
University of North Carolina at Charlotte
{*mshehab,mnko,htouati*}@*uncc.edu*

Abstract

Mapping user profiles across social network sites enables sharing and interactions between social networks, which enriches the social networking experience. Manual mapping for user profiles is a time consuming and tedious task. In addition profile mapping algorithms are inaccurate and are usually based on simple name or email string matching. In this paper, we propose a Game With A Purpose (GWAP) approach to solve the profile mapping problem. The proposed approach leverages the game appeal and social community to generate the profile mappings. We designed and implemented an online social networking game (*GameMapping*), the game is fun and is based on human verification. The game presents the players with some profiles information, and uses human computation and knowledge about the information being presented to map similar user profiles. The game was modeled using incomplete information game theory, and a proof of sequential equilibrium was provided. To test the effectiveness of the mapping technique and detection strategies, the game was implemented and deployed on Facebook, MySpace and Twitter and the experiments were performed on the real data collected from users playing the game.

1 Introduction

Social network (SN) services have been one of the main highlights of Web 2.0. Popular SN sites have attracted millions of users, for example Facebook hosts over 500 million users. Different SNs provide users with different sets of services and experiences, for example, Facebook and MySpace allow users to creates photo albums, fan clubs, and post feeds along with sharing all this content with friends, Twitter provides users with the ability to post short messages, and LinkedIn enables users to connect with other users for professional purposes. To enjoy these services, users endup creating accounts on different sites, for example most Twitter users have a Facebook account [14], and 64% of MySpace users have accounts in Facebook [28]. With the increasing popularity of SN connect services [18], this enabled users to connect websites with their SN accounts and to share their opinions and comments across networks. Leading SN sites are moving towards meeting the user's cross site interactions demands [20]. Users are able to connect different SN accounts and to share data across SNs, for instance, a user could connect his Twitter feed to his Facebook status such that his Facebook status will be updated automatically whenever he updates his Twitter feed [3].

When users create new accounts on a site they will spend time trying to rebuild their friendship connections with users they know, to alleviate this task several sites provide users with "import your friends" capability. For example, Bob has an established account in Facebook, and Bob heard from his friends about the video posting services provided by MySpace, so Bob creates a new account in MySpace which offers him to import his friends from Facebook. Using this functionality MySpace imports profile attributes of Bob's Facebook friends, and attempts to locate users who have similar attributes in MySpace (name, location, email hash, etc.) and recommends to Bob to add them as friends in MySpace. This approach is not effective in locating users with popular names, or for users who don't have matching attributes. Studies have shown that users tend to enter false information in their profiles [30], which causes attribute based matching approaches to generate inaccurate results [35, 39]. Furthermore, graph matching solutions are computationally expensive and require the knowledge of the complete graph of both networks [7, 8, 2]. Email based matching is only available when users use a same email across sites. A simple solution would be possible if all sites use a federated identity such as OpenID [31], however this technology is not popular among social network users.

In this paper, we propose a Game With A Purpose approach to solve the profile mapping problem. The proposed approach leverages the game appeal and social community to generate the profile mappings. We designed and implemented an online social networking game (*GameMapping*), the game is fun and is based on human verification. *GameMapping* takes advantage of people's existing perceptual abilities and desire to be entertained. The game will present the player with a user from one social network, and a set of friends from another social network, which represent the *set of mapping recommendations*. The friends' information is summarized in a profile card which includes the profile photo, name, age, location, etc. The player gets a small number of points for choosing one of the provided mappings, this reinforces a sense of *incremental individual success* in the game. The game also rewards *social success* by awarding the player a large number of bonus points when other users or friends agree to the player's provided mappings. This proposed mechanism is similar to social buying, where buyers are offered discounts discount deals (bonus) if they sign up for a deal in large masses [27]. Users will be allowed to invite their friends to play the game in hope of gaining the large bonus points. Similar games with a purpose have been successfully proposed to aid in labeling and tagging images over the web [33]. We also investigated several approaches for generating the set of mapping recommendations. The proposed *GameMapping* game was analyzed using game theory, to identify equilibrium under the current assumptions and point granting scheme to ensure that rational players will provide accurate profile mappings to maximize their game score. We performed several experiments to evaluate our approach on the game results, and we compared it to attribute based mapping which is presented in the experimental section. The main contributions of the paper are summarized as follows:

- We proposed a Game With A Purpose approach for solving the profile mapping problem as a game supported by social verification.
- We proved the equilibrium of the game scoring mechanism using game theory to ensure that rational players will provide accurate profile mappings while playing the game.
- We implemented our game as an online social networking game in Facebook, MySpace and Twitter. This implementation is a proof a concept and was used to collect and perform experimental results.

The rest of this paper is organized as follows. Section 2, provides an overview of Game With a Purpose and social networks. Section 3, defines the problem of profile mapping across sites. Section 4 describes how the proposed game works, and gives game details that include recommendation mechanism and the game theoretic proof. Section 5, describes the implementation of game system and the experimental results. The related work is discussed in Section 6, and Section 7 provides the paper conclusion.

2 Preliminaries

2.1 Game With a Purpose

Games with a Purpose (GWAP) is a form of human computation [33, 34], which gets humans to play enjoyable games that are also productive tools. These games are used in tasks that are hard for computers but easy for humans. For example, the ESP game [33] is a two-player game used for labeling and tagging images over the web, the game is setup to reward players providing the same labels by giving them bonus points if their tags match. Our goal is to design a GWAP to solve the profile mapping problem between social networks, by asking players to map their friends in the different social networks. One of the main challenges is the design of a points system that rewards correctly identified profile mappings and to maximize the reward for truthful rational players, and minimize the reward of irrational players. Gaming on social network platforms is becoming very popular with games such as FarmVille in Facebook [13] hosting over 62 million monthly active users. Our proposed game can easily be deployed on social network sites as an online game, and if it is popular we estimate that most of the account mappings can be properly discovered in a matter of weeks.

2.2 Social Networks

Users and relationships between users are the core components of social networks. Each user manages an online personal profile, which usually includes information such as the user's name, birth date, address, contact information, emails, education, interests, photos, music, videos, blogs, and many other items. Each user $u_i \in V$ maintains a profile P_i, which is composed of N profile attributes, $\{A_1^i, \ldots, A_N^i\}$. Each attribute is a name-value pair (an, av), where an and av represent name and value respectively. For example, a Facebook user profile includes attributes such as birthday, location, gender, religion, etc. Users are also able to post objects such as photos, videos, and notes to their profiles to share with other users.

Users are connected to a set of friends, using this notion a social network can be modeled as an undirected graph $G(V,E)$, where the set of vertices V is the set of users, and the set of edges E is the set of friendship relationships between users. The edge $(u_i, u_j) \in E$ implies that users u_i and u_j are friends. Using the graph based model for social networks, we leverage the node network structural properties to provide additional user attributes. These attributes include several small world network metrics such as: node degree centrality, betweenness, hit rate, eigen values [24]. For a user u_i, we are able to compute M network metrics $B_i = \{B_1^i, \ldots, B_M^i\}$. Each network attribute is similarly represented as a name-value pair (bn, bv) that will be added to the user personal profile attribute previously stated to constitute the user profile P. The neighborhood of user u is the subgraph $\mathcal{N}_u = (V_u, E_u)$, where $V_u = \{v | v \in V, (u,v) \in E\} \cup \{u\}$, $E_u = \{(x,y) | x,y \in V_u, (x,y) \in E\}$.

3 Problem Definition

The global profile mapping is defined as follows:

Definition 1 *(Profile Mapping Problem). Given social networks SN_A and SN_B, with social graphs $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$ respectively, find the set of profile mappings M of the form $(u_i, u_j) \in M$ where $u_i \in V_A$ and $u_j \in V_B$ belonging to the same user in both social graphs G_A and G_B.*

The problem of mapping data concepts between different sites or platforms have been applied to multiple areas, such as: database schema matching [21, 29], web search [10, 5], ontology mapping [9] and visualization [12, 38]. The graph isomorphism is an NP-Complete problem which involves finding one to one mappings between vertices and edges of a pair of graphs [4, 16]. The subgraph isomorphism graph matching problems has been proven to be NP-complete [15]. Furthermore, when $|V_A| \neq |V_B|$ known as the inexact graph matching problem, the complexity is proven in [1] to be NP-complete. In addition, the inexact sub-graph matching problem is NP-complete, and the largest common subgraph problem is also equivalent in complexity to the later which is NP-complete. Several attribute, model, object recognition and network based techniques were proposed to provide heuristic approaches to solving graph matching problems [7, 8, 2], these approaches are computationally expensive, and require the knowledge of the complete graphs G_A and G_B. In this paper, we propose solving the profile mapping problem by using human computation in the form of an online game.This approach has been used in [34, 33] to effectively map tags to images. The main assumption is that with the correct set of incentives, users would enjoy playing a game and at the same time contribute to mapping profiles between users in different networks.

Definition 2 *(Local Profile Mapping Problem) Given a user u who has identities u_i and u_j in social network SN_A and SN_B respectively, and user's local neighborhoods $\mathcal{N}_{u_i}^A$, $\mathcal{N}_{u_j}^B$ find the set of mappings $M_u \subseteq M$ mappings between profiles in $\mathcal{N}_{u_i}$ and $\mathcal{N}_{u_j}$.*

Our proposed approach will leverage the individual and social knowledge of social network users to provide mappings, and to provide mapping verifications which can be then used to solve the local profile mapping problem. The local profile mapping problem does not require knowledge of the whole social network graph, instead it only requires knowledge of the neighborhood network. Providing incentives to ensure the wide spread adoption of the game would allow solving a large number of local profile mappings, which enables the mapping of all similar profiles in large social networks. In fact, this is equivalent to the generalization of the subgraph isomorphism mappings of local networks to the maximum number of common subgraph problem in the global networks [40]. In this paper we are interested in studying mappings between social networks user accounts like Facebook, MySpace, and Twitter. Mapping profiles in social networks is applicable to identity management, and is a step towards enabling cross site interactions between users in different sites.

4 General Game Description

Our proposed game is called *GameMapping*. The basic idea is that players gain points by providing mappings of their friends' profiles in multiple social networks. *GameMapping* allows players to map Facebook and MySpace profiles, or Facebook and Twitter profiles.

In order to play the game the player needs to complete an authentication stage that involves two social network sites. We implement Facebook Connect, MySpaceID and TwitterID to enable the user to authenticate into the corresponding social networks and to authorize the GameMapping site access to the user's profile and friends list. This enables the GameMapping site to retrieve the user's profile and neighborhood social graph data which includes last name, first name, gender, age, country, profile picture, friends list and mutual friendships. This data enables our system to compute the local neighborhood for the current player $(\mathcal{N}_u^A, \mathcal{N}_u^B)$. The user profile referred to as the focus user u_f is presented to the player for mapping. The focus user u_f is selected from the neighborhood with the smaller size (cardinality). Without loss of generality we assume $|\mathcal{N}_u^A| \leq |\mathcal{N}_u^B|$, where the focus user profile u_f is selected from neighborhood $\mathcal{N}_u^A$. The game computes the recommended mappings profiles R from neighborhood $\mathcal{N}_u^B$ based on attribute and network distance metric. The focus user and the computed recommendations are then presented to the player. Figure 1, shows a screen shot of the game, where the focus user is in the center surrounded by his possible best recommended mappings displayed in a random order. The users' profile pictures are shown along with their profile information which include, age, gender, and location. Information about the recommended mappings is presented to the user when the mouse is moved over the photo. The player should decide either to map the focus user to one of the recommended profiles or to skip if no map is present. The player is given 40 seconds to make a decision about the presented game data set, then a new game data set is presented. The game also presents top 10 players ordered by the points earned. To motivate players into making cor-

Figure 1: The GameMapping Screen Shot

rect decisions of either mapping or skipping, the game awards the player 10 points for any provided map, 100 bonus points if the provided map is confirmed by another player, and 30 bonus points if a skip is confirmed by another player. In order to maximize the points (reward), a player should focus on providing the mappings that will most probably be confirmed by other players. When a player start the game, the player first plays the game with the player own network data set. In other words, the player maps friend's profiles. After the player is done mapping his local network, the player plays the game with a game dataset that is randomly selected. It ensures that players provide mappings towards multiple local profile mappings and at the same time ensure the game continuity. By motivating players to play multiple data sets enables the game to provide mapping confirmations as will be discussed in the game theoretic proof. In addition, each game dataset represents a local mapping problem, which when combined for multiple data sets results in the global mapping of the overall social network graph.

4.1 Recommendation Generation

Given a player u who owns profiles u_i and u_j, and the neighborhoods $\mathcal{N}_u^A$ and $\mathcal{N}_u^B$ the focus user u_f is selected randomly from the neighborhood that has the smaller number of nodes, which we refer to as the focus network. This design choice was made as the maximum number of possible mappings is equal to $min(|V_u^A|, |V_u^B|)$. Figure 2, shows both neighborhoods and the focus user u_f. Lets

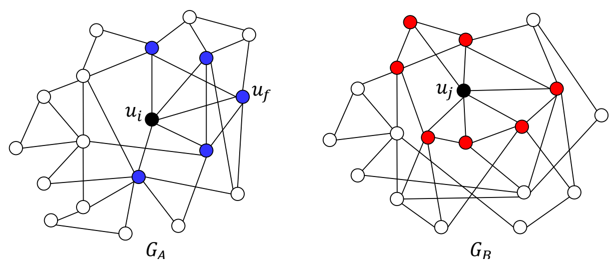


Figure 2: Neighborhood and Focus User Recommendations.

assume the focus user u_f is selected from $\mathcal{N}_u^A$. Given the focus user the mapping recommendation is generated by ranking the user profiles in $\mathcal{N}_u^B$ based on their similarity to the focus user. The similarity between two profiles is computed as a weighted sum of distances between the different user profile and network attributes. The profile attributes include first name, last name, gender, age and address. The network attributes include the centrality, betweenness, hit rate, degree and eigen values [6, 25]. We investigated several vector distances which include the Chebychev and Minkowski distance for numerical attributes, Cosine and Levenshtein distance for nominal attributes, and the Euclidian distance for the numerical attributes (i.e. age) and the Levenshtein distance for nominal attributes (i.e. gender, name) [19]. The weights of each attribute were computed based on a linear regression classifier trained using the knowledge collected from our initial experiments [36, 37]. The recommendation set R is the sorted list of proposed user profiles based on their computed similarity with the focus user. As indicated in Figure 1, the game presents the user with the top 12 recommended mappings select from the recommendation set R following the Top-k Fagin's algorithm [11]. The selected recommendations are shuffled randomly then displayed in a clock-wise fashion around the focus user. This randomization is required to ensure that players put some effort in finding the possible profile mapping among the displayed 12 recommendations. Moreover, by randomizing the recommendation set R this would avoid possible collusion between different players as each player is presented with the same 12 recommendations but not in the same location on the screen.

4.2 Game Theoretic Analysis

In this game the players do not communicate and each player does not know the action taken by other player. The game can be modeled as a two player extensive game with incomplete information. In this game the players are provided with a focus user u_f and a set of recommended mappings $R = \{u_1, \ldots, u_n, \phi\}$. Each player has a set of $n+1$ actions of the form $a_k = \mathbf{map}(u_f, u_k)$ where $u_k \in R$. Note, the action $a_{n+1} = \mathbf{map}(u_f, \phi)$, which is equivalent to the $\mathbf{skip}(u_f)$. The set of actions $A_1 = A_2 = A$, and the utility (δ_i) of player i is selected to satisfy the following conditions:

- $\delta_1 = \delta_2 = \delta$,
- $\delta(a_i, a_j) = \delta(a_j, a_i)$,
- $\delta(a_i, a_i) > \delta(a_i, a_j)$ for all $i \neq j$,
- $\delta(a_i, a_i) > \delta(a_{n+1}, a_{n+1})$ for all $1 \leq i \leq n$

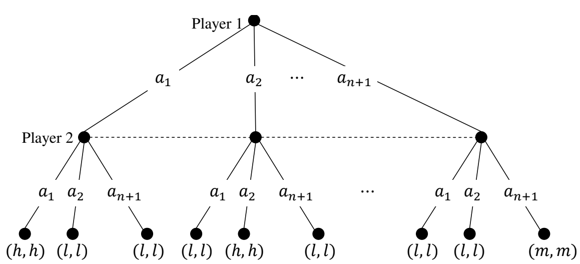


Figure 3: Game tree with imperfect information.

Figure 3, shows the extensive game tree, where each nodes represent players and edges represent player actions. The payoffs for players 1 and 2 are shown at the terminal nodes. The values of h and l are chosen such that $h > l$, this ensures that $u(a_i, a_i) > u(a_i, a_j)$ for all $i \neq j$. This game is a coordination game in which each player is trying to make the same choice as the other player to maximize their utility.

Rational players intend to maximize their expected game payoff. Note that the payoff from agreeing on a map is higher than the payoff from agreeing on a skip $(h > l)$, this motivates rational players to try to find possible maps between the focus user and one of the recommendations and to skip if they can not find a suitable map. The Nash equilibrium is a commonly used equilibrium notion that provides an equilibria such that no player can profitably deviate from and enhance their payoff with the belief that other players will not deviate [26]. Referring to the game representation in table form in Figure 4, The game has $n+1 = |A|$ pure Nash equilibria represented by the set S where $S = \{(a_i, a_i) : a_i \in A\}$, that is strategy that would result in maximizing the user payoff is when both users make the same action.

Player 1

Player 2	a_1	a_2	...	a_n	a_{n+1}
a_1	*(h,h)*	*(l,l)*	...	*(l,l)*	*(l,l)*
a_2	*(l,l)*	*(h,h)*	...	*(l,l)*	*(l,l)*
:	*(l,l)*	*(l,l)*	...	*(l,l)*	*(l,l)*
a_n	*(l,l)*	*(l,l)*	...	*(h,h)*	*(l,l)*
a_{n+1}	*(l,l)*	*(l,l)*	...	*(l,l)*	*(m,m)*

Figure 4: Game Nash Equilibria Indicated in Grey.

Since the game has multiple equilibria it is still not clear what action strategy with a rational player act upon. Given that each player does not know the action taken by the other player, the question that each player asks themselves is that given $\{u_f, R\}$ "what would other players do if they are presented with the same $\{u_f, R\}$?" and by the theory of focal points [22] players will usually coordinate at points that in some sense stick out from the others (focal points). A player game strategy can be described based on the probability of selecting an action a_i from the action set A given the focus user and recommendation set $\{u_f, R\}$. The probability $p(a_i|\{u_f, R\})$ represents the probability of choosing an action a_i conditioned on the game parameters $\{u_f, R\}$, which can be represented as $p(a_i|\{u_f, R\}) = p(a_i) \times r(a_i, \{u_f, R\})$. Where $r(a_i, \{u_f, R\}) = \frac{p(a_i, \{u_f, R\})}{p(a_i) \times p(\{u_f, R\})}$ is the relevance of action a_i to the set $\{u_f, R\}$. According to focal point analysis, a rational player would choose the action that maximizes the $p(a_i|\{u_f, R\})$ which is the action that is most relevant to the current $\{u_f, R\}$ set, which is described as follows:

$$a^* = \arg\max_{a_i \in A} p(a_i) \times r(a_i, \{u_f, R\})$$

By choosing action a^* players maximize their chance of being matched by other players in the system and ultimately gaining the payoff $\delta(a^*, a^*)$.

Assuming players are rational and they will choose the action that is most relevant for the given focus user and recommendation set, a dominant strategy that ensure that players coordinate and maximize their expected utility is attained when players follow the same actions selection probability $p(a_i|\{u_f, R\})$ [32]. This implies that players will be motivated to provide a map when they recognize a map and will prefer to choose skip if a map does not exist.

5 Experiments and Results

5.1 Implementation Details

We implemented the GameMapping game as an online game. The online game is functional on client browsers supporting Adobe Flash. The game communicates with a centralized GameMapping server to exchange and retrieve data. The game server is responsible for retrieving user profiles from social network sites, generating focus user and recommendation data sets, and storing all the mapping information. To support these features, we implemented social web application tools and APIs in the game server. Figure 5, depicts the architecture of our

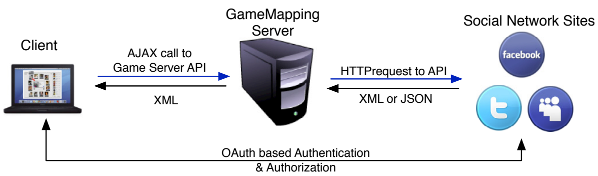


Figure 5: The Architecture of GameMapping

system. The game server connects to the each social network site using social web application tools such as Facebook Connect, MySpaceID, and TwitterID. These tools allow our game server to interact with the APIs of each social network site on behalf of game players. Facebook Connect is based on OAuth 2.0 specification while MySpaceID and TwitterID are based on OAuth 1.0a specification. We also implemented social plugins such as Like Button and Invitation to enhance the popularity and adoption of our game through friend of friend invitations and word of mouth. We implemented a polling mechanism to enable the retrieval of user's profile information, that is based on both server and client technologies (Ajax).

5.2 Collusion and Irrational Behavior

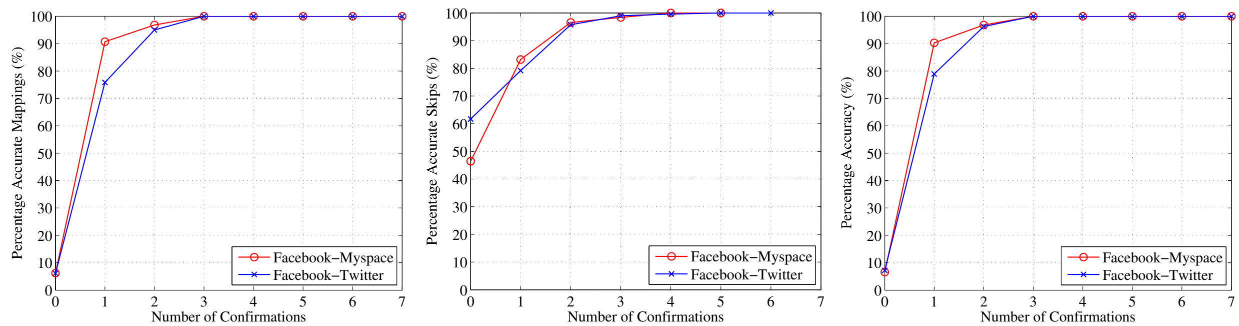
It is possible that some players map non-mapping and incorrect profiles intentionally. Based on the game theoretical discussion in Section 4.2, rational users are able to maximize their payoff by selecting the correct actions (map or skip). Irrational players are players who attempt to play the game and provide inaccurate mappings in hope of gaining high points or simply affecting our mapping accuracy. Although our game system does not provide a chatting feature, players might collude using another communication channel such as AIM or MSN chat, in order to provide the same inaccurate mappings to the game. To prevent collusion among players, our game displays randomly selected data sets to different players, who are allowed to play each game data set only once. Another irrational behavior is a player providing inaccurate mappings continuously by guessing, and getting l points for each provided map or skip. The game scoring mechanism ensures that rational players converge to a high score faster than guessing players.

In addition, we insert detection datasets into the normal game datasets to detect the irrational players. The detection game datasets are normal dataset that do not contain any correct mapping. If a player provides many mappings for the detection game dataset, there is high probability the player is an irrational player. We also recorded the amount of time taken by players in making each mapping to detect the irrational players and robots. If a player is an irrational player or a robot, the player might spend less time in each single mapping than rational players since the irrational players might provide mappings without comparing profiles. The game provides a CAPTCHA if the response rate is above the normal rate to prevent robots from playing the game. Finally we applied mapping confirmation strategy. If an irrational player provides inaccurate mappings, there is a low chance the inaccurate mapping gets a confirming map from other rational players.

5.3 Experiments

To evaluate our approach, we recruited participants which have accounts in multiple social networks by inviting users from MySpace and Twitter groups and apps on Facebook. As an incentive to play the game, we held a two week game competition to encourage users to participate in our research and distributed 10 iTunes gift cards to the top 10 players and an iPod Nano to the top player. One hundred and twenty-four players agreed to play the game, of which 80 where male, 32 female and 12 did not indicate their gender. There were two kinds of game the Facebook-MySpace (FB-MS) game for mapping user profiles between Facebook and MySpace and the Facebook-Twitter (FB-TW) game to map Facebook to Twitter. The FB-MS game was played by 30 players, and 94 players registered and played the FB-TW game. Perhaps users favored playing the FB-TW game due to the increasing popularity of both Facebook and Twitter. During the two weeks game competition, we collected 38,532 Facebook profiles, 8,452 MySpace profiles, 11,775 Twitter profiles and 7,411 profile mappings between user profiles. The collected profiles were used to generate the game datasets which were presented to the players to provide mappings between profiles in different networks. The game presented the players with a privacy consent that indicated that only the public information will be shared with other players which included the user's first name, last name, and location.

For verification and experimental purposes we manually verified all the provided profile mappings provided by the players using a simple verification web tool that shows details of mapped user's profiles with an inspection form. We designed the tool to generate comparison results of last name, first name, age, and gender automat-



(a) Mapping Accuracy (b) Skip Accuracy (c) Overall Accuracy

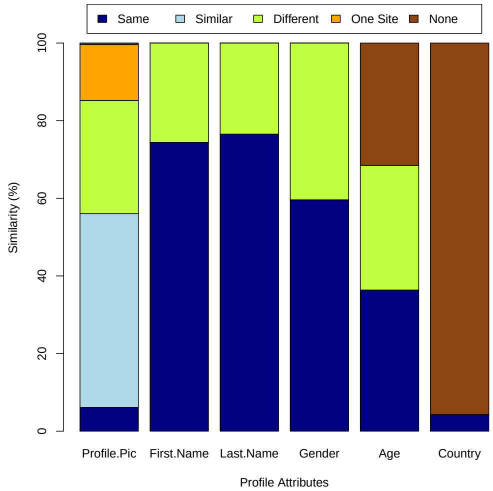
Figure 6: GameMapping Experimental Accuracy Results.

ically and we manually input the comparison result for profile pictures and countries. For each profile mapping we compared the profile pictures and categorized them into one of 5 types which include, Same, Similar, Different, Picture present only in one site, and None (picture is not present). In case of address and location information, geocoding distances were used to compare both profiles. If the profile information was not enough to make a decision, the inspectors visited profile page in each social network site to compare both profiles.

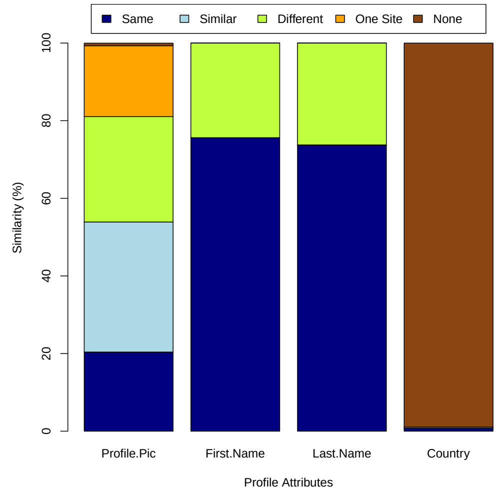
5.4 Evaluation of Mapping results

We analyzed the number of player confirmations required for accurate profile mappings and skippings by comparing the mappings provided by the players with the mappings verified manually. Figure 6(a) presents the mapping accuracy for different number of confirmations for both kinds of games (FB-MS and FB-TW), as shown the mapping accuracy increases as the number of confirmations increase. Note that, the mapping confirmation plateau's at 100% after 3 confirmations, which indicates that we need at least 3 confirmations to support 100% accuracy and 2 confirmations for 95% mapping accuracy. Figure 6(b) presents the skipping accuracy, which follows a similar pattern as the mapping accuracy as it also plateau's at 100% accuracy after 3 player confirmations for both FB-MS and FB-TW games. The FB-MS mapping and skipping results show a higher accuracy when compared to the FB-TW case, this is because the FB-MS dataset provides more user profile information to the player such as gender, age, address and other attributes that may help players in easily locating similar profiles accurately. Further, the friend relationship of Facebook and MySpace is based on mutual agreement and following relationship of Twitter is not based on mutual agreement. Therefore, the mutual agreement based relationship provides more knowledge for friends and higher accuracy. Figure 6(c) shows the over all confirmation accuracy for both the map and skip cases, which also plateau's at 3 confirmations.

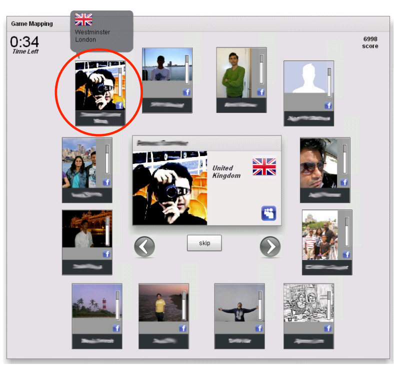
Figure 7(a) depicts the contribution of each profile attribute in verified FB-MS mapping results. Six attributes such as profile picture, first name, last name, gender, age, and country were used in comparing the profiles in the game. Note that, only 5.6% of users post exactly the same profile picture and 96.4% of users do not use a same profile picture (48.7% use similar pictures, 31.6% use different pictures, 13.7% have a profile picture in only one site, and 0.4% of the users do not have profile pictures). This shows that players mapped the same profiles based on other knowledge such as friendship information even if the two profiles did not use the same profile pictures. Last name and first name are important attributes in attribute based mapping, our results show that 74.4% of the users have the same last name and 72.8% users have the same first name. Which indicates that if the profile mapping is performed by comparing the name attributes, we expect about 73% matching accuracy. In other words, our game based mapping approach with confirmation is able to detect profile mappings for none matching profile names and provide a 27% improvement over the name based mapping. If gender and age are considered in attribute based mapping, the mapping result is not expected to increase as this usually missing or is low quality. Figure 7(b) depicts the contribution of each attribute in the verified FB-TW profile mapping results. In Twitter, only four attributes are used to compare the profiles in the game which include, profile picture, first name, last name, and country. The FB-TW attributes show a pattern similar to the FB-MS attributes. The minor difference is in the percentage of profiles that use the same profile pictures, last name and first name, where FB-TW shows higher percentages of similar profile attributes. The reason might be Facebook and Twitter are currently very popular sites. It makes many users to



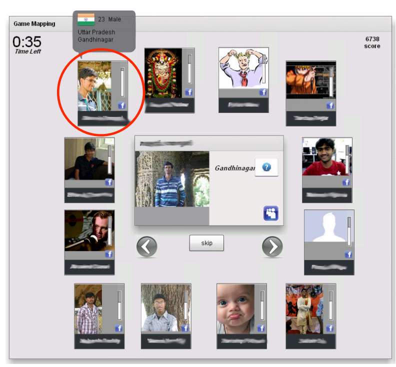
(a) Attribute Similarity in FB-MS



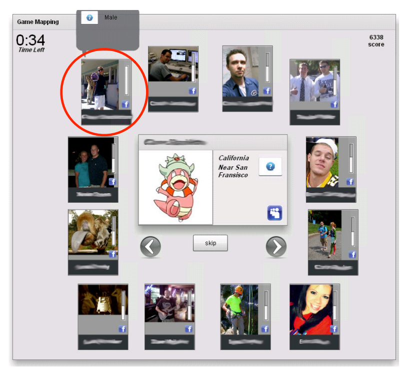
(b) Attribute Similarity in FB-TW



(c) Same Profile Photo



(d) Similar Profile Photo



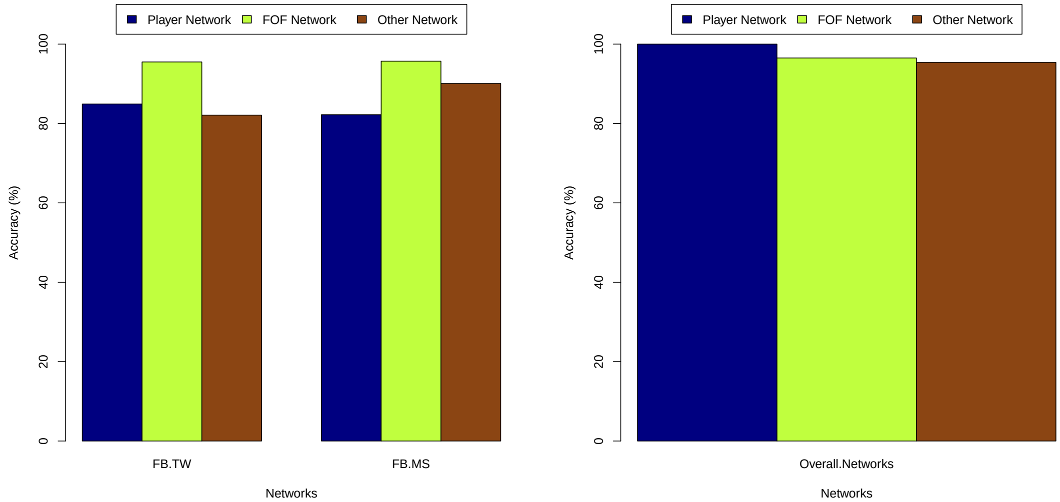
(e) Different Profile Photo

Figure 7: GameMapping Attribute and Photo Statistics

keep their profiles consistently up to date. In comparison to the name based attribute mapping, the FB-TW shows a 25% improvement in mapping accuracy. Figures 7(c)-7(e), show the possible profile mappings with respect to the same, similar and different profile photos, note that some users had the same, similar and different profile photos. The mapped user is indicated by the red circle.

To better understand how other network based approaches perform in matching the collected profile data. We used the similarity flooding graph matching approach [23], which matches profiles based on both profile attributes and network neighborhood similarity. The algorithm takes two labeled graphs (game data sets) as input and produces as output a mapping between matching profiles. We applied the collected game datasets to the similarity flooding algorithm and the generated an average matching accuracy of 74%. This result is far less than our proposed game mapping approach. The low accuracy generated by the similarity flooding approach could be attributed to the low similarity between the mapping neighborhoods which reduces the effectiveness of flooding algorithm. As indicated in Figure 7(a) and 7(b) profile attributes used in different social networks have a low degree of similarity, users do not always provide correct data or data is missing, attribute similarity is important in similarity flooding as it is used in initialization and flooding phases of the similarity flooding algorithm. In addition the neighborhood graph information for users in different social networks do not have considerable similarity in friendship connections and neighborhoods which tends to reduce the effectiveness of the flooding based similarity. On the other hand, our proposed approach provides higher accuracy due to the fact that player's map profiles not only based on the profile attributes but also based on the player's implicit knowledge about the profiles and on the reasoning behind of likelihood of mapping confirmation.

The game datasets are generated from the player's net-



(a) Accuracy and Network

(b) Accuracy of knowledgeable players

Figure 8: GameMapping Experimental Results.

work, Friend of Friend (FOF) network, and other user's network data. Figure 8(a) depicts the average accuracy of mapping results for different network types. For both FB-MS and FB-TW games, the results show that the accuracy of player network is lower than the accuracy of FOF network. The results did not meet our expectation that the accuracy of player network is higher than the accuracy of FOF network, which would be in turn higher than the accuracy of other network, since the players have more knowledge about their friends. We investigated the whole process of the game to answer the question why the accuracy of player network is lower than the accuracy of FOF network. First, we found that most players did not watch the video tutorial that is on the game homepage before they started the game. It made the players start the game without the knowledge about the game. Second, the players first played the game for their network dataset. Therefore, the players learned how to play the game while they were making incorrect or correct mappings on their network dataset. Then, they were able to play better when they played on the FOF network or other user's network game datasets. To confirm our discovered cause, we also investigated the mapping data. Figure 8(b) depicts the accuracy of knowledgeable players who knew how to play the game before starting the game. The knowledgeable players provided 100% accuracy on their network, 96.5% accuracy on FOF network, and 95.5% accuracy on other networks. It shows the players' friend relation influence on the accuracy of mapping results. The players provided higher accuracy on their friend profile mappings than unknown people's profile mappings. In summary, the game based profile mapping approach with confirmation provides better mapping results when compared to simple attribute mapping approaches. It is able to generate 100% accurate profile mappings with 3 or more mapping confirmations. Friend relation knowledge influences on the accuracy of mappings for different network types.

5.5 Irrational Player Detection Evaluation

In the initial stage of game design, we considered the irrational players and designed prevention and detection strategies as described in Section 5.2. To identify the irrational players, we calculated the mapping accuracy distribution of players as presented in Figure 9.

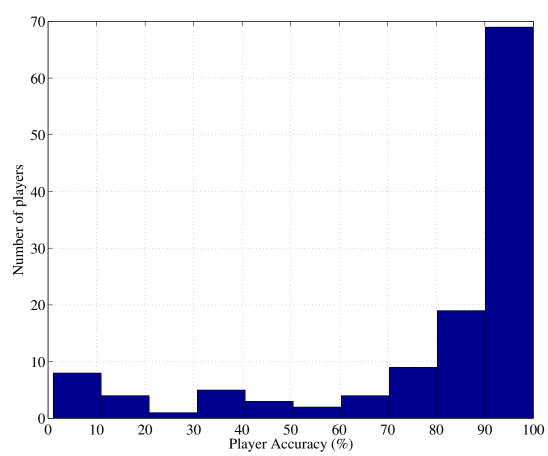


Figure 9: Accuracy Distribution of Players.

In our game period, 69 players provide over 90% mapping accuracy (18 players provided 100% mapping accuracy), and 8 players provided less than 10% mapping accuracy. We classify irrational players as either passive or active irrational players. A passive irrational player is a player that provides a small number of mapping which is lower than the average mapping of all the game players (105 mappings), and has an accuracy of 20% or less. On the other hand, an irrational player is considered active if he provides more than the average number of mappings and has an accuracy of 20% or less. Based on this classification, we discovered 12 irrational players, with 9 passive and 3 active irrational players. The passive irrational players provided 14 mappings on average, which implies that most passive irrational players did not spend much time in playing the game and left it shortly after their registration stage. There might be several reasons behind the reason for their low accuracy, one possible reason is that they did not understand the game and decided to test it out by providing random mappings. Table 1 shows a summary of the results extracted from the 3 active irrational players. The player 1 spent on average

Irrational Player	Mapping	Accuracy	Average Time
Player 1	130	6,15%	7.00 sec
Player 2	551	3.62%	0.55 sec
Player 3	2643	1.05%	1.65 sec

Table 1: Active Attackers

7 seconds to map each profile and provided 130 mappings with 6.15% accuracy and the player 2 spent 0.55 seconds to map each profile and provided 551 mappings with 3.62% accuracy. Both players have low accuracy but it is evident that player 2 did not review the focus user data or the recommend user profiles instead he preferred to randomly map or skip the presented user. All the three players did play the detection game, and all of them provided 0% mapping and skipping accuracy for the detection game. Therefore, all the above 3 players were detected by the detection game strategy. Another detection strategy was based on comparing the average mapping time, where the average mapping time of the players who have accuracy above 90% was 6.7 seconds. On the other hand, the average mapping time for the irrational players was 3 seconds. This implies that rational players spend about twice the time to map profiles when compared to irrational players. Moreover, most mapping results from the irrational players did not get a confirmation, and they were not in the top 10 players.

6 Related Work

Mapping users account across social networks is an important task that will allow users and third party applications to interact across social networks. In this paper, we divided our literature review to the following areas: attribute matching, graph matching, and human computation using games.

Without a common identity management system between different sites, attribute matching techniques are used to detect the same user in different sites by utilizing user's information. Wang et al. [35] proposed a record comparison algorithm that detects deceptive criminal identities using four personal attributes: name, date of birth, social security number and address. It calculates the overall similarity score of personal attributes. If the overall similarity score is higher than a pre-defined threshold, two people are considered a matched people. The authors also revealed that incomplete records with many missing data could significantly increase the error rate of the record comparison algorithm that is a common limitation of many identity matching techniques using only personal attributes. Jennifer et al. [39] showed that combining social features with personal features could improve the performance of criminal identity matching. They artificially constructed incomplete datasets from a complete datasets by randomly choosing a percentage of person's records and removing their data of birthday or address values. Using this incomplete dataset with a decision tree classification method, they found out if the dataset had more missing values in personal identity attribute, the social contextual features significantly increase the matching performance. This paper showed how personal attributes and social features affect the performance of the identity matching.

The graph matching problem was classified as one of the most difficult problems. In fact, many categories of graphs were classified as NP-compete problem in [16]. Exact subgraph matching problem, for example, where the number of vertices in each subgraph is the same was proven to be NP-complete by [15], however under certain constraints, where the subgraph is a tree in the big forest graph, it was proven to be resolved in polynomial time. In our paper, we consider the inexact subgraph matching problem, where the number of vertices (nodes) in each network subgraph is different, and this problem was also proved to be NP-complete by [1]. In [23] the other use a directed graph matching approach for database schema matching consisting of similarity flooding with a fixed point computation of similarity. In our paper we represent the social networks with undirected graphs.

Using human knowledge for computation while entertaining them is one of the increasing trends in the recent years. Most of the research applications of this technique is in the image labeling problem that is described in [33], where the authors created an image labeling game called the ESP Game to take advantage of the powerful vision sense and common knowledge of humans to achieve the labeling. The ESP game is played by two players without any information or link between each other but the image being labeled, and they are asked to label objects that are present in the image. Once the players agree on an object that is present in the image they will be introduced with another image and so on. Another good game that used human common knowledge for semantic annotation is PhotoSlap [17]. In PhotoSlap, the authors based their idea on the ESP game and the popular Snap card game, where the players flip cards containing random images, and *slap* each time they identify two consecutive images of the same person. In addition, the game supports the *objection* and *trap* actions to enforce truthfulness, where the players are presented with a set of images that they can set as *traps* (i.e. photos containing similar faces/heads) at the beginning of the game. Once a player *slaps*, the other players may *object* to the truthfulness of the *slap*, which is verified by the *traps* defined earlier in the game. Our idea is similar to the ESP and PhotoSlap games in the way of using human knowledge to map between user accounts using not only images, but also profile attributes, such as: age, gender, first name, last name and other attributes that might be helpful for a human to make a mapping decision.

7 Conclusion

In this paper, we presented the Game With A Purpose (GWAP) approach that solves the profile mapping problem. We provide two type of games: Facebook-MySpace (FB-MS) game and Facebook-Twitter (FB-TW) game. To detect irrational player who provide incorrect mapping intensionally, we also designed and applied an irrational player detection strategies to our game system. In our experiments, the proposed detection strategies detected irrational players effectively. It discovers the active irrational player spent 50% less time than rational players for mapping and their most mapping results did not get the agreement from other players. The evaluation of mapping results show our proposed mapping approach generate higher mapping accuracy (FB-MS: 27% improvement, FB-TW: 25% improvement) than the name based mapping results. We also observed that users are able to accurately map their friends, friend of friend and other network profiles. Finally, we showed that accurate mappings can be concluded if 3 or more rational players agree on it. In the future, we will extend this work to support other social networking sites, and to deploy the game on these sites.

8 Acknowledgments

This research was partially supported by grants from the National Science Foundation (NSF-CNS-0831360, NSF-CNS-1117411) and a Google Research Award.

References

[1] ABDULKADER, A. M. Parallel algorithms for labelled graph matching. *PhD thesis, Colorado School of Mines.* (1998).

[2] AUWATANAMONGKOL, S. Inexact graph matching using a genetic algorithm for image recognition. *Pattern Recognition Letters 28*, 12 (2007), 1428 – 1437.

[3] B. SCHWARTZ. How to connect twitter to facebook status updates, `http://www.ehow.com/how_4668396_connect-twitter-facebook-status-updates.html`, 2010.

[4] BASIN, D. A. A term equality problem equivalent to graph isomorphism. *Inf. Process. Lett. 51*, 2 (1994).

[5] BLONDEL, V. D., GAJARDO, A., HEYMANS, M., SENELLART, P., AND DOOREN, P. V. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM* (2004).

[6] BORGATTI, S. P., AND EVERETT, M. G. A graph-theoretic perspective on centrality. *Social Networks 28*, 4 (October 2006), 466–484.

[7] CESAR, R., BENGOETXEA, E., AND BLOCH, I. Inexact graph matching using stochastic optimization techniques for facial feature recognition. *Pattern Recognition, International Conference on 2* (2002), 20465.

[8] CESAR, JR., R. M., BENGOETXEA, E., BLOCH, I., AND LARRA NAGA, P. Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recogn. 38*, 11 (2005), 2099–2113.

[9] DOAN, A., MADHAVAN, J., DOMINGOS, P., AND HALEVY, A. Ontology matching: A machine learning approach. In *Handbook on Ontologies in Information Systems* (2004), Springer-Verlag.

[10] DONG, X., HALEVY, A., MADHAVAN, J., NEMES, E., AND ZHANG, J. Similarity search for web services. *VLDB* (2004).

[11] FAGIN, R., LOTEM, A., AND NAOR, M. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences 66* (2002).

[12] FAN, K.-C., LU, J.-M., AND CHEN, G.-D. A feature point clustering approach to the recognition of form documents. *Pattern Recognition 31*, 9 (1998).

[13] FARMVILLE GAME. Zynga game network inc., `http://www.facebook.com/FarmVille`, 2010.

[14] FRANKY BRANCKAUTE. Twitter's Meteoric Rise Compared to Facebook [Infographic]. The Blog Herald, June 2010.

[15] GAREY, M. R., AND JOHNSON, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, January 1979.

[16] GAREY, M. R., AND JOHNSON, D. S. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1990.

[17] HO, C., HSIANG, T., AND HSU, J. Photoslap: A multi-player online game for semantic anotation. *AAAI'07* (2007).

[18] KO, M. N., CHEEK, G., SHEHAB, M., AND SANDHU., R. Social-networks connect services. *IEEE Computer 43*, 8 (aug. 2010), 37 –43.

[19] LIU, B. *Web DataMining Exploring Hyperlinks, Contents, and Usage Data*. Springer-Verlag, 2007.

[20] M. GUMMELT. Publishing to twitter from facebook pages, `http://blog.facebook.com/blog.php?post=123006872130`, 2010.

[21] MADHAVAN, J., BERNSTEIN, P., CHEN, K., HALEVY, A., AND SHENOY, P. Corpus-based schema matching. In *In ICDE* (2003), pp. 57–68.

[22] MCADAMS, R. H. A focal point theory of expressive law. *Virginia Law Review 86*, 8 (2000), pp. 1649–1729.

[23] MELNIK, S., GARCIA-MOLINA, H., AND RAHM, E. Similarity flooding: A versatile graph matching algorithm. In *Ontology handbook* (2002), pp. 117–128.

[24] NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E 64*, 1 (June 2001).

[25] NEWMAN, M. E. J. The structure and function of complex networks. *SIAM Review 45*, 2 (2003), 167–256.

[26] OSBORNE, M. J., AND RUBINSTEIN, A. *A Course in Game Theory*. The MIT Press, July 1994.

[27] P. BOUTIN. Investors hot on social shopping: First groupon, now livingsocial, `http://venturebeat.com/2010/04/29/livingsocial-funding`, 2007.

[28] PATRIQUIN, A. Connecting the social graph: Member overlap at opensocial and facebook, . `http://blog.compete.com/2007/11/12/`, Nov 2007.

[29] RAHM, E., AND BERNSTEIN, P. A. A survey of approaches to automatic schema matching. *VLDB* (2001).

[30] REALWIRE. Social networking sites: Almost two thirds of users enter false information to protect identity, `http://www.realwire.com/release_detail.asp?ReleaseID=6671`, 2007.

[31] RECORDON, D., AND REED, D. Openid 2.0: a platform for user-centric identity management. In *DIM '06: Proceedings of the second ACM workshop on Digital identity management* (NY, USA, 2006), pp. 11–16.

[32] SUGDEN, R. A theory of focal points. *Economic Journal 105*, 430 (May 1995), 533–50.

[33] VON AHN, L., AND DABBISH, L. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems* (NY, USA, 2004), pp. 319–326.

[34] VON AHN, L., LIU, R., AND BLUM, M. Peekaboom: a game for locating objects in images. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (NY, USA, 2006).

[35] WANG, G. A., CHEN, H., XU, J. J., AND ATABAKHSH, H. Automatically detecting criminal identity deception: An adaptive detection algorithm. *IEEE Trans. on Systems, Man and Cybernetics, Part A* (2005).

[36] WITTEN, I. H., AND FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2 ed. Morgan Kaufmann, 2005.

[37] WITTEN, I. H., AND FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*, second ed. Morgan Kaufmann, 2005.

[38] WU, H., CHEN, Q., AND YACHIDA, M. Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 6 (1999), 557–563.

[39] XU, J., WANG, G. A., LI, J., AND CHAU, M. Complex problem solving: identity matching based on social contextual information. *Journal of the Association for Information Systems, Vol 8, Issue 10* (Oct 2007).

[40] ZAGER, L. *Graph Similarity and Matching*. Master dissertation, MIT, 2005.