Design of Social Games for Collecting Reliable Semantic Annotations

Lilian Weng*, Rossano Schifanella† and Filippo Menczer*
*School of Informatics and Computing, Indiana University, Bloomington, USA
†Department of Computer Science, University of Turin, Italy

***Abstract*—We present two social tagging games based on the *Chain Model* for object association. *GiveALink Slider* and *Great Minds Think Alike* harness human power to generate large streams of social tagging data. Such social annotations are utilized to help people organize Web resources and infer semantic relationship, which in turn can enhance Web applications such as search, recommendation, navigation, and categorization. The two games leverage several design features as well as external social media resources to create entertaining incentives for the players to tag sites and multimedia objects. Preliminary analysis of data generated by these games suggest that the proposed model can be effective for the collection of annotations that are novel, have high quality, and lead to reliable semantics.**

I. Introduction

Social tagging provides us with a powerful paradigm to manage online resources collaboratively. Users can freely choose words to describe resources and therefore resources get descriptions from various users represented as sets of tags, collectively forming a *folksonomy.* Tags are widely used in many Web 2.0 and social media Web sites, by which users can organize their own collections, discover interesting resources, and find friends with similar interests. For instance, in Delicious, Flickr, YouTube, and Last.fm, users tag their links, photos, videos, and music, respectively; in Twitter, hashtags have emerged as a de-facto convention to assign conversations to informal topical channels. These folksonomies, generated by the power of the crowd, not only help enrich the semantic space of online resources, but also enhance the performance of many Web services, such as search and recommendation.

Users share annotations largely for their own individual needs and aspirations, sometime leading to low-quality annotation data. In current social tagging systems, there is little motivation for the majority of users to annotate many resources with sufficient numbers of accurate tags, and the number of new pages that are posted per day to social annotation systems is small compared to the rate of growth of the Web [9]. This causes a significantly sparse semantic network of social annotations. Without any control on tagging behaviors, users can easily employ poor tags or even abuse the system by spamming [13]. They can use tags that are unrelated to a resource, too general to meaningfully describe a given resource, or so specific that they are only useful for one individual.

In this paper, we adopt the *Game With A Purpose* (GWAP) [1] approach to accelerate the generation of reliable social annotation data for supporting other useful Web services. The use of GWAP to engage humans in the solution of hard computational problems has gained popularity in recent years in the field of Human Computation [3]. Two social tagging games, *GiveALink Slider* and *Great Minds Think Alike*, proposed here aim to generate large streams of high-quality social annotations as a side effect of enjoyable activities.

The development of games to help enhance both the quantity and quality of annotation data is an integral component of the `GiveALink.org` project, which broadly examines several aspects of social tagging with the goal of fostering the construction and applications of socially driven semantic annotation networks. Both of the games introduced here are built upon existing work in the GiveALink project. Previous research includes the design of effective similarity relationships among pages, tags, or users [14], [20], applications to page recommendation [16], exploratory navigation interfaces [8], bookmark management [19], and social spam detection [13]. We wish to explore the use of this prior work, especially the *Maximum Information Path* (MIP) similarity measure [15], in our design of effective tagging games.

We present relevant background and related work in the next section. Then in § III we outline a general model for object association games. We discuss details of game design in § IV, and a preliminary evaluation in § V.

II. Background

In a social tagging system, an *annotation*, also known as *triple*, is defined as a tripartite relationship (u, r, t) between a *user* u, a *resource* r, and a *tag* t. A *post* is a set of triples $(u, r, \{t_1 \cdots t_k\})$ sharing a user and a resource, or a relationship between a user, a resource, and a *set of tags*. In the remainder of the paper we will also refer to a *link* as a relationship (r, t) between a resource and a tag, supported by the users annotating the resource with the tag. Folksonomic collections of social annotation data have been shown to be useful to improve social navigation [17], Web search [9], [6], [24], personalized recommendation services [16], and social links prediction [20], [5].

The original purpose of games is amusement and recreation, but it is still possible to use entertainment to guide players to other objectives with specific rules and designs. *Serious games* are games designed in the context of training in military, education, and public health settings [25]. They are used to incentivize the process of learning with interesting game components. It is reasonable to assume that when people are highly involved they are more likely to learn more and better.

GWAP are designed in a way that players can help solve hard computational problems while having fun. GWAP are also defined as *crowdsourcing through play*. Crowdsourcing aims to harness human knowledge from a community to accomplish tasks hard for machines but simple for humans. Crowdsourcing techniques have been utilized in many areas, including general human knowledge repositories such as the Wikipedia, human subject task marketplaces such as the Amazon Mechanic Turk [11], and scholarly data collection and impact evaluation as done via the Scholarometer tool [10].

The best known instance of GWAP is the ESP Game [2], which takes on the challenge of image recognition by asking pairs of players to reach an agreement on image labels. Verbosity [4] works in a similar way, but it collects commonsense knowledge for helping build intelligent systems. The Great Minds game was partly inspired by an online word association game called *Human Brain Cloud*, which generated interactive visualizations of crowdsourced word networks. In prior work we have described an early prototype of a social tagging game, which was a precursor of the Slider game presented here [22]. The preliminary evidence collected from the annotations produced by that game suggests the need for incentives against overly general tags, which is one of the mechanism design principles behind the model presented here.

Games are also used for scientific and other purposes. Foldit is a game for non-expert players to handle computationally difficult protein folding problems in biochemistry [7]. Galaxy Zoo (galaxyzoo.org) leverages human image recognition abilities to classify celestial objects. Microsoft uses games in Club Bing (clubbing.com) to promote the Bing search engine. In all the club games, players have to use Bing as a tool to collect enough hints for winning the game and therefore get a chance to try some functions that they might not have noticed before.

Social games are games associated with multiple players. The interactions among players increases the entertainment of the game. Browser-based social games are attracting attention recently, as they make it easy to gather a large number of players without a requirement of certain software or monetary costs [12], [21]. Mobile games are another hot topic based on the human desire to play wherever and whenever a user is.

III. Design Principles

Both tagging games presented here are based on the general idea of building a chain of semantically related objects. The objects are connected based on a measure of similarity, and the players extend the chain by making these relationships explicit. The idea is formalized as the *chain model* [23] for object association games that collect descriptions about, and discover hidden relationships among, Web resources, media, people, and geographical locations.

The chain model consists of an ordered sequence of objects $\langle obj_0, \ldots, obj_n \rangle$, and the last element obj_n is called the *active* object. Chain model games allow players to characterize an object obj_i with a set of descriptions $D^i = \{desc^i_0, \ldots, desc^i_{k_i}\}$ in some language. At each step the player p can add a new description to obj_n or make a game *move* to extend the chain, i.e., a transition $obj_n \overset{\mu}{\mapsto} obj_{n+1}$ where μ is a user-defined relation. To support the player in the decision of the next object in the chain, the model suggests a set of *candidate* objects C^n that are computed from a system-defined measure of similarity with respect to the active object.

While playing, users can connect to like-minded people who describe objects or build the chain in a similar way, with social connections created or reinforced as a result. We also incorporate some common game design features to assure enjoyability, such as score rankings and rewards for accomplishing specific tasks. The output of the chain model is a stream of posts $\langle (p, obj_i, D^i) \rangle$ where $0 \leq i \leq n$, representing how a player with handle p describes an object obj_i. We set up two mechanisms to guarantee the relevance of the player's description with respect to the object: *internal* and *external* verifications. The former checks multiple user agreement on each object-description association; the latter confirms the relations through trusted external resources. Differences between browser and mobile platforms are considered in the game designs as well. All of these features characterize games based on the chain model as a platform to implement social games with the purpose of collecting reliable object description data.

IV. Tagging Game Design

Both of our tagging games, built upon the proposed chain model, are aimed at collecting annotations of Web resources with trusted descriptive tags. To connect related Web pages or tags in the chain, both games need an effective measure of similarity among objects. To this end we employ a scalable, collaborative measure of similarity for social annotation systems, named Maximum Information Path (MIP) [15]. MIP similarity has been implemented and integrated into the GiveALink system, and used as an important component in several applications, such as recommendation, bookmark management, and spam detection. MIP similarities among pages or tags are available via the public GiveALink API (GiveALink.org/api_doc).

A. Browser Game: GiveALink Slider

In the GiveALink Slider, the chain objects are Web pages. Players build chains of pages and generate descriptions by tagging these pages. The origin page obj_0 in the chain is provided randomly by the game, and then the player starts to extend the chain by tagging the first page with one or more relevant tags. Each time the player enters tags for a page, the game displays a small set of pages based on the player's tags as candidates for the next object in the chain. The player is free to choose any candidate as the next page to tag. Only the active (newly added) page can be tagged, but all the previous tags can be easily reviewed. The player may win badges by completing predefined tasks (§ IV-A4).

1) Interface: Slider has four interface modules: *Play, Visualization, History,* and *Rank.* In the *Play* space (Fig. 1), the page chain is presented in a carousel, and the player can switch among pages. The URL, title and thumbnail of each page help the player try relevant tags. Old pages have their tags shown

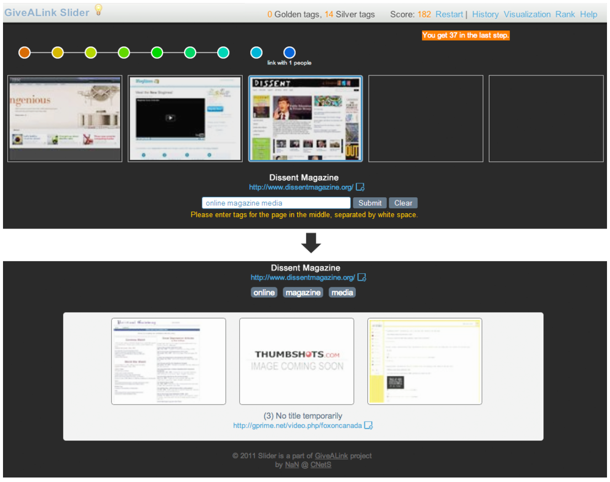


Fig. 1: Slider interface for tagging and selecting pages.

when selected, while the active page has an input field for new tags. Once new tags are entered, the player can choose among three candidate pages displayed below the chain. The chain is also displayed as a series of colored nodes: connected adjacent nodes represent related pages.

2) Mechanisms: The annotation data generated in Slider is expected to be merged into the GiveALink triple store to enrich the semantics network of annotations. Therefore, the game is designed to strengthen relationships that seem to be weaker than they should be, or discover new relationships. The game mechanisms incorporate two important design principles from the chain model: (i) Players gain points for relevant tags, and lose points when the chain gets disconnected because of unrelated or unverified tags. (ii) Links between tags and resources that exist in the GiveALink triple store are trusted, so they can be used as benchmarks. New links are recorded until verified by multiple players internally.

Each user-submitted tag for a given page (each link) can be classified into one of three types:

- ***used:*** tags that have been used for this page in GiveALink;
- ***suggested:*** tags that have been suggested for this page by previous players, but not in GiveALink;
- ***new:*** tags that have not been used or suggested for this page by GiveALink users nor other players.

Because *used* and *suggested* tags are verified by GiveALink or game players, they are more reliable and they are worth more points. *New* tags may or may not be relevant, so we need to wait for further evidence; they may turn into suggested tags when they are used for the same page more than once. Suggested tags may turn into used tags if they are confirmed by enough players.

Once a player tags the active page, candidate pages are presented to help him expand the chain (Fig. 1). The candidates are resources related to the active page to keep the chain connected. However, if none of the newly entered tags can be verified, then random candidate pages are displayed and the chain gets disconnected. The candidate page selected by the player becomes the new active page to be tagged, and so on. During each step, the player can be connected to other players if they share the same links (Fig. 2). Visualizations of the whole relationship network among players who share common tags or pages can also be viewed.

Additionally, we have designed special features to make the game easy to play and therefore fun. First, origin pages are not selected completely at random. Since the player gains points by providing verified tags, two factors directly influence the difficulty: (i) whether the player is familiar with the page, and (ii) whether previous players have tagged this page. Slider selects a starting page that was previously used in the game with some probability (30%). Otherwise, it obtains a page from the GiveALink API. If the player has shared a sufficiently large number of pages (100 or more) in the GiveALink system, the API returns one of the pages in her personal collection with some probability (50%), or else it picks a random page among all resources (20% probability).

A second design feature allows players to discard pages that are spam, broken links, or written in an unknown language, since those pages distract players and result in poor annotations. The player can replace a candidate page by reporting the reason. All the spam or broken link reports are stored, flagging suspicious or invalid resources and helping improve the accuracy of the GiveALink spam detector [13].

3) Scoring: The score of each tagging step is obtained by adding scores associated with each tag. Each tag is worth a number of points that depends on its *specificity* and *type*. Specific, trusted, and novel tags are most valued. The score for each tag t for the active page is given by two terms: $s(t) = \eta(t) + \lambda(t)$, where η is a measure of tag *specificity* and λ is a function of trustworthiness and novelty. The specificity is calculated as: $\eta(t) = \epsilon / \sum_{t' \neq t} \sigma(t', t)$, where σ is the MIP similarity function between a pair tags [15] and ϵ is a parameter. Tags similar to many other tags are general and therefore are worth fewer points.

According to the tag classification, the trustworthiness and novelty functions λ are defined as follows:

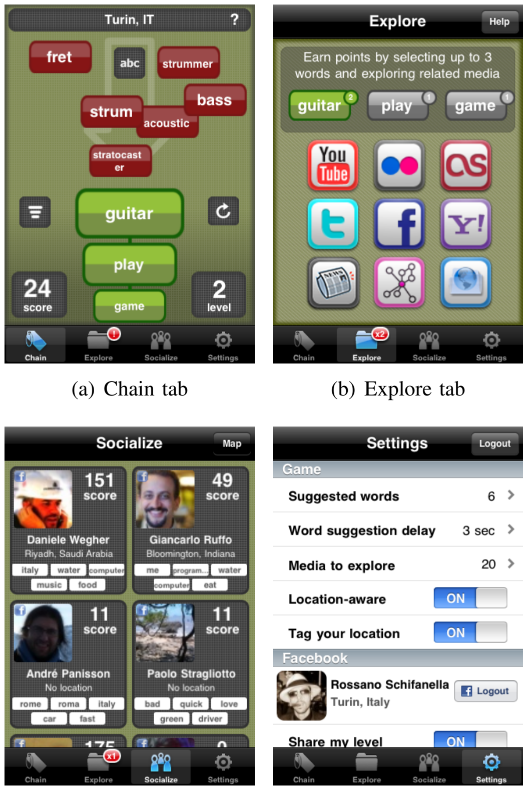
- Discover trusted links: $\lambda(t_s) = \alpha \cdot f(t_s)$, where t_s is a suggested tag, f is the number of players who have suggested t_s for the active page, and α is a parameter. The player gets more points when more people have agreed with him in previous games.
- Find related tags: $\lambda(t_u) = \beta$, where t_u is a used tag for the active page and β is a parameter. We thus reward players who use relevant tags that have previously been used to annotate the same resource by GiveALink users.
- Save untrusted tags: $\lambda(t_n) = \gamma$, where t_n is a new tag for the active page and γ is a comparatively small parameter.

Scores are summed across tags in a post during each game step, and accumulated across the steps, yielding a cumulative score. We set the parameters such that $\alpha > \beta > \gamma$ to prioritize based on the trustworthiness and novelty of posts: among trusted posts, novel ones (*suggested*) are more valued than known ones (*used*); untrusted (*new*) posts are valued the least.

TABLE I: List of Badges in GiveALink Slider.

Type	Description
	Chain with some tagged pages
	Linked with some people in a single game
	Linked with some people
	Single game with score above a threshold
	Total score above a threshold
	Connections between some pages

Fig. 2: Visualization of Slider social network: when player *konata* tags soic.indiana.edu with *indiana*, *research* and *informatics*, she finds that player *narcopimp* has tagged the same page with *research* and *indiana* as well, and has also used *indiana* for another page (www.iub.edu).



(a) Chain tab (b) Explore tab

(c) Socialize tab (d) Settings tab

Fig. 3: Great Minds main operational modes.

4) Enjoyability: To make the game more enjoyable, there are six types of *badges* corresponding to different game tasks, and each has multiple levels of difficulty (Table I). The player accomplishes these tasks to win badges and corresponding bonuses. The tasks deal with pages in the chain, connections with other players, and score milestones. The harder the task, the greater the bonus. We display both the score and the numbers of different badges for each player in the rank section.

Another incentive comes from visualizations of *social connections*. There are two types of ties among players. A *strong connection* occurs when two players tag the same page with the same tag — they generate the same link. A *weak connection* is established between players by either common tags or common Web pages across all games. The game visualizes social ego-networks with both types of ties (Fig. 2).

5) Output Quality: If we reward every new tag, it is easy for players to cheat by tagging pages with any unrelated terms. To prevent this behavior, Slider computes the proportion of trusted tags in each tagging step. A player is suspected of cheating if the proportion is too small, and she will lose some points as penalty. More specifically, a tagging step is deemed to be an instance of cheating if $\overline{\lambda(t)} - \gamma$ is smaller than a threshold, where the bar denotes average across tags in a post. When the player is punished, the candidate pages are chosen at random rather than among pages related to the active object, and consequently the chain gets disconnected.

The cold-start problem can negatively affect Slider output when few annotations are available to verify links. In our initial implementation we employ the Delicious API (delicious.com/help/api), and in particular the `posts/suggest` method, as an external resource to obtain recommended tags and accelerate the confirmation new triples. *New* and *suggested* tags are promoted to *used* and recorded in the GiveALink database if confirmed either by enough players or by Delicious.

B. iPhone Game: Great Minds Think Alike

Great Minds Think Alike is a word association game that lets the player build semantic concept networks and explore similarity relations between people, tags, and media content. The player builds a chain of words by entering or selecting related terms among suggested options coming from the GiveALink knowledge base. The player gains points typing new words or exploring content from social media sites like Flickr, YouTube, Last.fm, Twitter, and others. While exploring content, the game sends annotations to the GiveALink database through its public API. Tags in the chain are geo-tagged and players are linked to their Facebook profiles so they can find like-minded people nearby.

1) Interface: The game interface is composed of a tab bar controller at the bottom and a contextual view covering the rest of the screen. The controller allows the user to choose between different modes of operation by selecting the corresponding icon and displaying the view for that mode. Great Minds provides four functional modes: *Chain*, *Explore*, *Socialize*, and *Settings* tabs (Fig. 3).

TABLE II: Great Minds chain interface buttons.

Button	Name	Description
	keyboard	Enter a word by typing
	history	View/edit the current chain
	new	Start a new chain

The *Chain* tab (Fig. 3(a)) contains the main game screen where users build the sequence of semantically related words. At the top, a label indicates the geo-location of the player and a question mark leads to a help section. The connected green boxes display the last three words of the current chain and an animated arrow on the background indicates the *active* word. The user can select the next word in the chain by tapping one of the red terms that appear in random positions in the upper part of the screen. They fade out after a while leaving the stage empty for new suggestions. The *keyboard*, *history*, and *new* buttons shown in Table II let the player type a new word, edit the current chain, or start a new one, respectively. The player's score and level are shown as well.

The *Explore* tab (Fig. 3(b)) enables the exploration of media content related to the words in the chain. The player formulates a query by selecting up to three terms from the entire chain. Each word is displayed with a bonus. The user searches different types of content related to the query by tapping one of the social media buttons.

The *Socialize* tab (Fig. 3(c)) lists nearby users with similar associations to the player. Finally, a hierarchical table controller presents game preferences in the *Settings* tab (Fig. 3(d)).

2) Mechanisms: The main purpose of the game is to collect annotations from a set of social media sources that are considered reliable. This collection is carried out by the navigation of media content exposed through public Web services and it is based on the *assumption* that a player agrees on the resource-tag links when he consumes the social content resource by tapping on it, generating corresponding posts. The word association game thus drives search and annotation.

According to the proposed model, a Great Minds *chain* is an ordered sequence of words $\langle w_0, \ldots, w_n \rangle$. For the origin, players can *type a word* or start from a *random word* or the *Word of the Day*. Random words are chosen from the set of terms previously used in the game. Words that have a popularity below a threshold are filtered out to remove very rare and spam terms. The threshold is computed at runtime according to the frequencies of terms in the vocabulary. The Word of the Day is obtained from the *Wordnik* Web service (www.wordnik.com/developers).

A *move* $w_n \overset{\mu}{\mapsto} w_{n+1}$ is the selection of a *target* term that the player feels is related in some way to the *active* word. The target can be selected by entering a word from the keypad (see the *keyboard* button in Table II) or choosing one of the candidate terms that appear on the screen. These candidates come from two different sources: (i) the GiveALink knowledge base through the API method `Tag.getSimilar`, or (ii) the Great Minds moves dataset. In the latter case, given the active word w_n, the game server extracts all the moves involving w_n and ranks them by frequency — how many times the words have been associated in the past. Both sources are invoked in turn, depending on a rate parameter. This scheme increases both the heterogeneity and the reliability of the candidate suggestion mechanism; a source with an empty result set implies an automatic call to the other.

When a move μ is performed, the game stores on the server a set of metadata that enables geo-social features (§ IV-B4). Each move triggers a *record* (μ, id, lat, lon, ts) where id is the player's GiveALink email, lat and lon are the geo-coordinates of the player's current position, and ts is a timestamp. For privacy reasons the geo-coordinates are rounded to a lower precision so that, instead of the exact location, the system records a nearby point. The user can disable the geo-localization engine.

The *History* button (see Table II) allows the player to inspect the complete chain, backtrack a wrong move, or remove typos that are frequent using the keypad.

In the *Explore* screen the player navigates the Web accumulating points and generating annotations. The game provides several *channels* with content from a large spectrum of media types. We currently employ 9 APIs: YouTube for videos (code.google.com/apis/youtube), Flickr for images (flickr.com/services/api), Last.fm for music (last.fm/api), Twitter for microblogging (apiwiki.twitter.com), Facebook for social networking (developers.facebook.com), Yahoo! for Web pages (developer.yahoo.com), Google for news (code.google.com/apis/newssearch), Bibsonomy for scholarly papers (bibsonomy.org/help/doc/api.html), and GiveALink.

The user defines the query by selecting (green color) or deselecting (gray color) words in the chain. Navigating a *channel* yields a list of resources that have been tagged or indexed with all the terms in the query. The tap gesture on a result element reflects the user interest in that result and an implicit agreement on the resource-tag link. A detailed view of the resource is then displayed.

The player can control the behavior of the Great Minds game via three sections in the *Settings* panel. The *Game* section controls parameters related to the word suggestion scheme, the exploration of media content, and the localization. The *Facebook* section is intended to manage social networking preferences, and the *About* section contains additional information and links to online help resources. Particular emphasis is placed on the privacy settings:

- ***Location-aware:*** enables the geo-localization of the player's device. The current coordinates are sent to the game server when a new move is performed.
- ***Tag your location:*** if the location-aware option is on, the game allows the player to tag their current location in the form of a Google map URL. This annotation is sent to the GiveALink server along with other posts related to media content.
- ***Share my level:*** this option controls the post on the player's Facebook wall when a new game level is reached.

3) Scoring: Great Minds allows a player to gains points by (i) extending the chain with a move or (ii) exploring social media in the descriptive phase. In the former case, the game rewards the player with one point when the target is specified by typing a new word. The idea is to foster the introduction of serendipitous associations and fresh thoughts that enrich the game knowledge base.

In the latter case, the systems rewards the generation of trusted annotations: the more triples, the more points. Each query word is assigned a *bonus* in the interval $[1, 3]$ that represents the value of the term in points. As an incentive for novel associations, the bonus associated with a target word w_{n+1} is inversely related to the popularity of the corresponding move μ. Let $f(\mu)$ be the number of times μ was taken by other players. We define two thresholds $\tau_1 < \tau_2$ and assign 3 points if $f(\mu) < \tau_1$, 2 points if $\tau_1 \leq f(\mu) \leq \tau_2$, and 1 point if $f(\mu) > \tau_2$. In the current setting $\tau_1 = 2$ and $\tau_2 = 5$, but thresholds can be updated according to the distribution of move frequencies. For each media content explored by the player, the bonus is the sum of the single bonuses associated with the query words. The points earned are accumulated and shown in a badge over the *Chain* tab.

4) Enjoyability: To foster competition between players and make the game enjoyable, we associate *expertise levels* with cumulative scores. The level is defined such that players progress indefinitely over time, but novice players progress more quickly. With the player's permission, Great Minds posts an update on her Facebook wall when the score reaches a new level. This is designed to promote adoption of the game.

Geo-social interactions are another way to make the game more enjoyable. To enable the social features a player has to login with a Facebook account and allow the game to access the profile data through the Facebook Graph API (developers.facebook.com/docs/api). Great Minds does not manage or store credentials since the login procedure is performed directly by the Facebook servers through the OAuth protocol (auth.net). The game stores the name, email, location, and profile picture of a user. Facebook data is used to provide the user with an option to discover other players, with whom she may wish to socialize through the social network. When a player enters the *Socialize* tab, the system shows similar users ranked according to a geographic proximity metric.

We define the similarity σ between players as a measure of their common moves. The geographic distance between players p_1 and p_2 is computed as the shortest distance between their locations according to the *orthodromic* definition: $\delta(p_1, p_2) = \phi \arccos(\cos \rho_1 \cos \rho_2 \cos(\omega_1 - \omega_2) + \sin \rho_1 \sin \rho_2)$ where ϕ is the mean Earth radius, and ρ_i and ω_i are the latitude and longitude of p_i's last move. Let us further define the function $getSim(p, \kappa_1, \kappa_2)$ where $\kappa_1 \geq \kappa_2$ are two system parameters. In a nutshell, the function extracts the κ_1 players most similar to p (according to the σ measure) and sorts them by their distance from p. The nearest κ_2 players are shown in the *Socialize* tab view.

5) Output Quality: As discussed in § III, the relevance of the player's description of an object is verified externally in Great Minds since it is implicit in the trust that the user places in the quality of the social media content and metadata. Although spam or low quality annotations are still possible, we are confident that the phenomenon is strongly mitigated. In any case, we are planning to introduce a consensus scheme similar to the approach adopted by the Slider game.

To mitigate the cold-start effect due to a new or rare tag, the game knowledge base is initialized with the concept network originating from the *Human Brain Cloud* dataset. This enables the suggestion feature in many cases in which the active term does not have direct contributions from the players.

Finally, the use of the *Word of the Day* service is intended to increase the probability of common moves among players and, consequently, the likelihood of geo-social connections and interactions.

V. Evaluation

The games presented here are designed with the purpose of generating high-quality social annotation data, therefore, it is critically important to evaluate how fast and how well the purpose is achieved. We evaluate the output of both games from several perspectives, to show that the games can collect reliable, novel triples comparatively fast, serving as a good complement to existing social tagging system.

A. Basic Statistics

The Slider game was open to the public through a 6-week online user study, and the Great Minds game has been available in Apple App Store for about one month. During this time, we were able to collect more than 40,000 triples from approximately 200 users. In Table III we can see that the chains built in both games have a reasonable length, and each object receives multiple descriptions. The game mechanisms allow users to contribute a large amount of triples through several runs of game play.

TABLE III: Basic game statistics.

	Slider		Great Minds	
	Mean	St. Dev.	Mean	St. Dev.
Descriptions per object	6.1	11.1	3	3.5
Objects per chain	25.9	67.0	7.2	16.5
Chains per player	5.3	13.4	24	44
Triples per player	260	834	134	206

B. Novelty

The social annotation generated by the games are significantly different from the existing data in social tagging systems. Among all tag-url links confirmed by multiple players in Slider, 5.74% pairs overlap with the GiveALink repository and 21.87% appear through the Delicious tag recommendation API. The overlap is even smaller in Great Minds: only 1.5% tag-url links exist in GiveALink. The majority of the tag-url links are novel associations discovered through the games.

Based on the triples collected in Slider and Great Minds, we calculate the MIP similarity [15] between each pair of tag. It is a number between 0.0 and 1.0, and the value is proportional

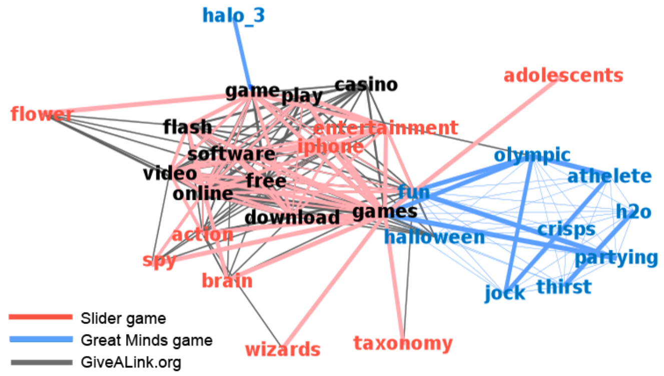


Fig. 4: Networks of tags most related to *games* in GiveALink (grey), Slider (red), and Great Minds (blue). The width of each edge is proportional to the similarity value.

to how related the tags are; a non-zero similarity indicates that two tags are related. We consider a pair of tags with non-zero similarity value in a game but zero similarity in GiveALink as a *novel* relation, i.e., a tag-tag relation that was previously undetected and is newly discovered through a game. Using this definition, 60.5% of the tag-tag relations in Slider and 52.9% of the relations in Great Minds are novel. These results suggest that the games are efficient at generating streams of novel social annotations.

C. Quality

In Slider, tags used to describe the same Web resource are considered as being related. In Great Minds, related tag candidates are generated on GiveALink (including game) data, but players reinforce tag-tag relations by constructing tag chains. With our additional design mechanisms for description verification, we expect the extracted relations among tags to be of good quality and specificity.

As an example, let us extract the top 10 tags that are most related to the tag *games* from the GiveALink system as well as both games. We visualize the similarity network among these tags in Fig. 4, where each tag is represented as a node, and each related tag pair is connected with an edge with width proportional to the similarity value. Most tags displayed are indeed well related to the topic “games.” This example highlights a significant difference between the annotations from games and GiveALink. Tags collected from the games are more specific, compared to several quite general terms in GiveALink, such as *online*, *free*, and *software*.

D. Density of the Semantic Space

With more game data being added to the semantic space, we expect the structure of the folksonomy built from the tripartite relationships between tags, users, and resources to show some changes. Since both games can collect novel descriptions and discover novel relations (see § V-B), we want to examine the similarity networks among tags and see whether they become denser and more “metric” by closing triangles [18].

To explore this question, let us construct networks of sampled tags from each game. Then we compare two different networks with the same set of tags, one with only relations in GiveALink and another with additional relations from a game. Table IV reports on several measures from such networks, based on samples of 2,000 tags from Slider and 410 tags from Great Minds. We clearly see that with game annotations, the tag similarity networks become denser and more metric, with fewer isolated small components and more closed triangles.

TABLE IV: Semantic network comparisons.

	Slider		Great Minds	
	GiveALink	Combined	GiveALink	Combined
Nodes (Tags)	2,000	2,000	410	410
Edges	53,443	57,533	8,387	12,208
Components	689	357	55	13
Clustering Coeff.	0.55	0.60	0.69	0.73
Closed Triangles	1,176,864	1,253,348	110,644	227,008

VI. Conclusions and Future Work

User-generated annotations are valuable for many Web services, such as providing a better Web search experience, or personalized recommendations. However, in current social tagging systems, users can easily employ poor tags or even abuse the system by tagging spam. The lack of control on social tagging yields tags that are overly general or personal, as well as spam. We propose games with the purpose of improving both quantity and quality of social annotations data.

We designed and developed two social tagging games, the GiveALink Slider on browsers and Great Minds Think Alike on mobile devices. The basic design principles are generalized into the chain model for games that generate object descriptions by object association moves. In our games, players win points by extending a chain of related objects, or providing descriptions for existing chain objects. A combination of a player, an object and a description is verified internally or externally to preserve the correctness of the data collected. Novel and uncommon tags are rewarded more to encourage fresh descriptive information.

Our evaluation of preliminary data from both games is very encouraging in regard to the novelty and quality of game annotation. Both game collections only have a small overlap with an existing social tagging system, and we are able to discover novel relations among tags. The quality of the data seems to be sustained by the game mechanisms. The output difference between our games and other social tagging systems originates from the primary goals of the two applications. Social tagging sites mainly work as online link repositories, allowing users to upload and share bookmarks. The games, on the other hand, ask players to tag or link resources directly. Game players tag for fun while thoughts about the objects flow freely in their mind, rather than for future individual reference. As a consequence, we are able to collect meaningful, descriptive relations and avoid overly general tags. The semantic networks of concepts that emerge from these annotations are dense and connected.

As the game data grows, we plan to conduct a more thorough evaluation and study several related research questions.

First, we would like to compare the novel relations from games with WordNet or similar resources to see whether they agree with each other. The results can be used as another evaluations of the quality of the game output. User behavioral patterns are another interesting topic. Since the game players must be registered users of GiveALink.org, it is likely that they have shared some bookmarks in the system. Then it is possible to compare the tagging behaviors of users in games versus their traditional social tagging behaviors: whether tags for a page used in a game differ from those from bookmarks; whether players prefer to use more specific or more general tags; and whether descriptions vary significantly across players given the same game goal. The results may help understand the reasons behind the small overlap observed between games and social tagging systems. Finally, the assumption in Great Minds that a tap gesture on a result media reflects an implicit agreement on the resource-tag link requires empirical validation; if it is not confirmed, the design will have to be tweaked to guarantee meaningful annotations.

Acknowledgments We are grateful to A. Panisson and S. Patil for assistance in the game implementation, K. Gabler for sharing *Human Brain Cloud* data, and the members of the Networks and agents Network (cnets.indiana.edu/groups/nan) at the IU School of Informatics and Computing for helpful suggestions during game design and testing. This work is funded in part by NSF award IIS-0811994 *Social Integration of Semantic Annotation Networks for Web Applications* and PRIN project *Information Dynamics in Complex Data Structures* of the Italian MIUR.

REFERENCES

[1] L. V. Ahn. Games with a purpose. *IEEE Computer*, 39(6):92–94, 2006.
[2] L. V. Ahn and L. Dabbish. Labeling images with a computer game. In *Proc. SIGCHI Conf. on Human factors in computing systems*, 2004.
[3] L. V. Ahn and L. Dabbish. Designing games with a purpose. *Communication of the ACM*, 51(8):58–67, 2008.
[4] L. V. Ahn, M. Kedia, and M. Blum. Verbosity: a game for collecting common-sense facts. In *Proc. SIGCHI Conf. on Human factors in computing systems*, 2006.
[5] L. M. Aiello, A. Barrat, C. Cattuto, G. Ruffo, and R. Schifanella. Link creation and profile alignment in the aNobii social network. In *Proc. Second IEEE International Conference on Social Computing (SocialCom)*, 2010.
[6] S. Bao, X. Wu, B. Fei, G. Xue, Z. Su, and Y. Yu. Optimizing web search using social annotation. In *Proc. 16th Intl. World Wide Web Conference*, 2007.
[7] S. Cooper, A. Treuille, J. Barbero, A. Leaver-Fay, K. Tuite, F. Khatib, A. C. Snyder, M. Beenen, D. Salesin, D. Baker, and Z. Popović. The challenge of designing scientific discovery games. In *Proc. Fifth Intl. Conf. on the Foundations of Digital Games*, 2010.
[8] J. Donaldson, M. Conover, B. Markines, H. Roinestad, and F. Menczer. Visualizing social links in exploratory search. In *Proc. 19th ACM Conf. on Hypertext and Hypermedia (HT)*, 2008.
[9] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *Proc. 1st ACM Intl. Conf. on Web Search and Data Mining (WSDM)*, 2008.
[10] D. T. Hoang, J. Kaur, and F. Menczer. Crowdsourcing scholarly data. In *Proc. Web Science Conference*, 2010.
[11] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proc. 26th annual SIGCHI Conf. on Human factors in computing systems*, 2008.
[12] C. Klimmt, H. Schmid, and J. Orthmann. Exploring the enjoyment of playing browser games. *CyberPsychology & Behavior*, 12(2):231–234, 2009.
[13] B. Markines, C. Cattuto, and F. Menczer. Social spam detection. In *Proc. 5th Intl. Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, 2009.
[14] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *Proc. 18th Intl. World Wide Web Conference (WWW)*, 2009.
[15] B. Markines and F. Menczer. A scalable, collaborative similarity measure for social annotation systems. In *Proc. 20th ACM Conf. on Hypertext and Hypermedia (HT)*, 2009.
[16] B. Markines, L. Stoilova, and F. Menczer. Social bookmarks for collaborative search and recommendation. In *Proc. 21st National Conf. on Artificial Intelligence (AAAI)*, pages 1375–1380. AAAI Press, 2006.
[17] D. Millen and J. Feinberg. Using social tagging to improve social navigation. In *AH2006 Workshop on Social navigation and community-based adaptation*, 2006.
[18] L. M. Rocha. Semi-metric behavior in document networks and its application to recommendation systems. In V. Loia, editor, *Soft Computing Agents: A New Perspective for Dynamic Information Systems*, International Series Frontiers in Artificial Intelligence and Applications, pages 137–163. IOS, 2002.
[19] H. Roinestad, J. Burgoon, B. Markines, and F. Menczer. Incentives for social annotation. In *Proc. 32nd Annual ACM SIGIR Intl. Conf. on Research and Development in Information Retrieval*, page 838, 2009.
[20] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer. Folks in folksonomies: Social link prediction from shared metadata. In *Proc. 3rd ACM Intl. Conf. on Web Search and Data Mining (WSDM)*, 2010.
[21] J.-M. Vanhatupa. Browser games: The new frontier of social gaming. *Communications in Computer and Information Science*, 84:349–355, 2010.
[22] L. Weng and F. Menczer. Givealink tagging game: An incentive for social annotation. In *Proc. ACM SIGKDD Human Computation Workshop (HComp)*, 2010.
[23] L. Weng, R. Schifanella, and F. Menczer. The chain model for social tagging game design. In *Proc. 6th Intl. Conf. on the Foundations of Digital Games (FDG)*, 2011.
[24] Y. Yanbe, A. Jatowt, S. Nakamura, K. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *Proc. 7th ACM/IEEE-CS Joint Conf. on Digital libraries*, pages 107–116, 2007.
[25] M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.

Lilian Weng is a PhD student in the School of Informatics and Computing at Indiana University Bloomington. She holds a bachelor in Information Systems and Computer Science from Peking University, China. She works as research assistant in the Center for Complex Networks and Systems Research. Her research focuses on social media, social games, human-computer interaction, and complex information networks modeling.

Rossano Schifanella is an assistant professor of computer science, and a member of the Security and Networking Group at the Universitá degli Studi di Torino, Department of Computer Science. He holds a Laurea and a Ph.D. in Computer Science from the Universitá degli Studi di Torino, Italy. His research focuses on distributed systems, mobile computing, web mining, socio-technical systems, modeling and visualization of complex information networks.

Filippo Menczer is a professor of informatics and computer science at Indiana University, Bloomington. He holds a Laurea in Physics from the University of Rome and a Ph.D. in Computer Science and Cognitive Science from the University of California, San Diego. His research focuses on Web, text, and data mining, social Web applications, distributed and intelligent Web information systems, and modeling of complex information networks.