

MOTIVATIONS FOR PARTICIPATION IN SOCIALLY NETWORKED COLLECTIVE INTELLIGENCE SYSTEMS

Jon Chamberlain, Udo Kruschwitz, Massimo Poesio
 School of Computer Science and Electronic Engineering
 University of Essex
 Wivenhoe Park, Colchester CO4 3SQ, England
 {jchamb,udo,poesio}@essex.ac.uk

ABSTRACT

One of the most significant challenges facing systems of collective intelligence is how to encourage participation on the scale required to produce high quality data. This paper details ongoing work with *Phrase Detectives*, an online game-with-a-purpose deployed on Facebook, and investigates user motivations for participation in social network gaming where the wisdom of crowds produces useful data.

INTRODUCTION

Recent advances in human language technology have been made possible by acquiring large-scale resources created by novel methods of harnessing collective intelligence over the Internet. This approach is motivated by the observation that a group of individuals can contribute to a collective solution, which has a better performance and is more robust than an individual's solution. This is demonstrated in simulations of collective behaviour in self-organizing systems (Johnson, Rasmussen, Joslyn, Rocha, Smith & Kantor 1998).

Collective Intelligence (CI) systems such as Wikipedia and similar large initiatives have shown that a surprising number of individuals can be willing to participate in projects.

A novel approach was the development of games-with-a-purpose (GWAP) that aggregate data from non-expert players for collective decisions similar to what might be expected from an expert.

Encouraging participation in GWAP projects remains a significant challenge, despite successful early efforts (von Ahn 2006).

This paper investigates user participation in CI systems and whether social network platforms could offer anything to this approach. The paper then goes on to describe *Phrase Detectives*, a GWAP for creating annotated language resources, specifically looking at the interface that was developed for the Facebook platform and the modifications that were made to maximise the social incentives for players. Data from over a year of the game being live on Facebook is analysed and, in conclusion, suggestions are proposed for developers considering using a social networking platform for deployment of CI interfaces.

COLLECTIVE INTELLIGENCE SYSTEMS

Collective intelligence can be shown in many domains including Computer Science, Economics and Biology¹ but here we focus on coordinating collective action in computational systems of CI. Individual decisions made by the community are aggregated in an attempt to produce a high quality, collective decision comparable to an expert judgement.

One important goal of CI systems is to overcome the bottleneck in creating and maintaining resources that would normally have to be done by paid administrators. Examples include encyclopedia websites like Wikipedia, citizen science projects for common knowledge and games-with-a-purpose that collect metadata.

By collecting decisions from a large, distributed group of contributors it is possible to approximate a single expert's judgements (Snow, O'Connor, Jurafsky & Ng 2008). This is in line with findings in other contexts (Feng, Besana & Zajac 2009, Albakour, Kruschwitz & Lucas 2010).

Participation

The willingness of Web users to collaborate in the creation of resources is clearly illustrated by Wikipedia. English Wikipedia numbers (as of October 2011) 3,773,941 articles, written by over 15.5 million collaborators and 5,559 reviewers.²

Citizen science projects, where non-expert volunteers complete large-scale or computationally difficult tasks, include Open Mind Commonsense³ (now ConceptNet⁴) which demonstrated that Web collaboration can be relied on to create resources, with 14,500 volunteers contributing nearly 700,000 sentences (Singh 2002).

The first, and perhaps most successful, game-with-a-purpose was The ESP Game⁵ which attracted over 200,000 players who have produced over 50 million labels (von Ahn 2006).

¹<http://scripts.mit.edu/~cci/HCI>

²http://meta.wikimedia.org/wiki/List_of_Wikipedias

³<http://openmind.media.mit.edu>

⁴<http://conceptnet.media.mit.edu>

⁵<http://www.gwap.com/gwap>



Figure 1. Screenshot of the Phrase Detectives Facebook homepage.

Clearly there is huge potential for the general public to become engaged in CI systems and collaborate in producing resources that would not be possible to achieve using traditional methods.

Social Networking Platforms

Given the social nature of CI it seems logical to deploy CI systems on platforms where the users are already networked. In recent years social networking has become the dominant pastime online. As much as 22% of time online is spent on social networks like Facebook, MySpace, Bebo, Twitter and others. This is three times the amount of time spent emailing and 7 times the amount of time spent searching the Internet.⁶

The success of social network games such as Cityville, with over 50 million active players each month, or The Sims, Farmville and Texas HoldEm Poker, with over 30 million active monthly players each, show that the potential for large scale participation is possible using social networking platforms.⁷

An estimated 927 million hours are spent each month by

Facebook users playing games⁸, which is another indicator of the vast human resource available.

A study of US and UK social network users showed that Facebook was by far the most frequently used platform for social network gaming (used by 83% of users, compared to MySpace, the next highest platform, at 24%).⁹ However Google are planning to increase their market share of this valuable resource.¹⁰

Human language technology games integrated into social networking sites such as Sentiment Quiz¹¹ on Facebook show that social interaction within a game environment does motivate players to participate (Rafelsberger & Scharl 2009).

It is becoming more apparent that CI interfaces should be linked to social networking sites like Facebook to achieve high visibility, to explore different ways players can collaborate and to exploit this enormous human resource.

⁶<http://mashable.com/2010/08/02/stats-time-spent-online>

⁷<http://www.appdata.com>

⁸<http://www.allfacebook.com/facebook-games-statistics-2010-09>

⁹http://www.infosolutionsgroup.com/2010_PopCap_Social_Gaming_Research_Results.pdf

¹⁰<http://www.reuters.com/article/2011/08/12/us-google-games-idUSTRE77A66H20110812>

¹¹<http://apps.facebook.com/sentiment-quiz>

Rhinogradentia (Wikipedia)

Rhinogradentia (also known as snouters or Rhinogrades or Nasobames) is a fictitious mammal order documented by the equally fictitious German naturalist Harald Stumpke. The order's most remarkable characteristic was the Nasorium, an organ derived from the ancestral species's nose, which had variously evolved to fulfill every conceivable function.

Both the animals and the scientist were allegedly creations of Gerolf Steiner, a zoology professor at the University of Karlsruhe. A mock taxidermy of a certain Snouter can be seen at the Musée zoologique in Strasbourg.

The order's remarkable variety was the natural outcome of evolution acting over millions of years in the isolated Hi-yi-yi islands in the Pacific Ocean.

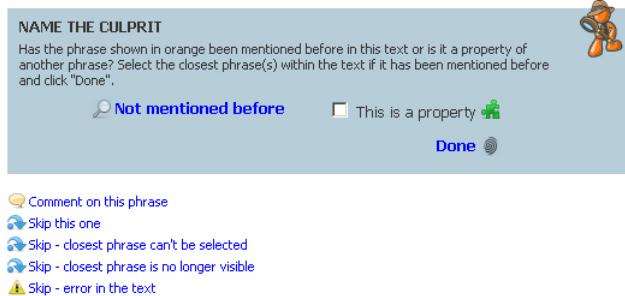


Figure 2. Detail of a task presented in Annotation Mode.

PHRASE DETECTIVES FACEBOOK INTERFACE

The first interface for the *Phrase Detectives* game¹² (Chamberlain, Poesio & Kruschwitz 2008) enabled groups of players to work on the same task over a period of time as this was likely to lead to a collectively intelligent decision (Surowiecki 2005).

The Facebook version of *Phrase Detectives*¹³, launched in February 2011, maintained the overall game architecture whilst incorporating a number of new features developed specifically for the social network platform (see Figure 1).

The game was developed in PHP SDK (a Facebook API language allowing access to user data, friend lists, wall posting etc) and integrates seamlessly within the Facebook site. Data generated from this version of the game is compatible with previous versions and both current implementations of the game run simultaneously on the same corpus of documents.

In order to play the game a Facebook user must grant certain permissions: the basic access (user details and friends list), which is required for all applications, and access to posting on the user's wall. Once the user has allowed the game access they never need to login to the game, only to Facebook.

The game uses 2 styles of text annotation for players to complete a linguistic task. Initially text is presented in Annotation Mode (called Name the Culprit in the game - see Figure 2). This is a straightforward annotation mode where the player makes an annotation decision about a

¹²<http://www.phrasedetectives.com>

¹³<http://apps.facebook.com/phrasedetectives>

Rhinogradentia (Wikipedia)

Rhinogradentia (also known as snouters or Rhinogrades or Nasobames) is a fictitious mammal order documented by the equally fictitious German naturalist Harald Stumpke. The order's most remarkable characteristic was the Nasorium, an organ derived from the ancestral species's nose, which had variously evolved to fulfill every conceivable function.

Both the animals and the scientist were allegedly creations of Gerolf Steiner, a zoology professor at the University of Karlsruhe. A mock taxidermy of a certain Snouter can be seen at the Musée zoologique in Strasbourg.

The order's remarkable variety was the natural outcome of evolution acting over millions of years in the isolated Hi-yi-yi islands in the Pacific Ocean.



Figure 3. Detail of a task presented in Validation Mode.

highlighted markable (section of text). If different players enter different interpretations for a markable then each interpretation is presented to more players in Validation Mode (called Detectives Conference in the game - see Figure 3). The players in Validation Mode have to agree or disagree with the interpretation.

Players are trained with training texts created from a gold standard (a text that has been annotated by a linguistic annotation expert). A player always receives a training text when they first start the game and may also need to complete one when being promoted to the next level. Once the player has completed all of the training tasks they are given a rating (the percentage of correct decisions out of the total number of training tasks). The rating is recorded with every future annotation that the player makes as the rating is likely to change over time.

The scoring system is designed to reward effort and motivate high quality decisions by awarding points for retrospective collaboration (see Figure 4).

The game makes full use of socially motivating factors inherent in the Facebook platform. Any of the player's friends who are playing the game form the player's team, which is visible in the left hand menu. Whenever a player's decision agrees with a team member they score additional points.

Player levels have well-defined criteria and the player must activate the new level once the criteria are met (see Figure 5):

- Total points scored
- The player's rating
- Documents completed
- Training documents completed
- Facebook posts made from the game



Figure 4. Detail of the reward screen, displayed at the end of each case, showing the player how many points they scored and who they agreed with.

The Facebook game has monthly and all-time leaderboards as well as leaderboards for the highest level players, highest rated players and the players with the biggest team.

INCENTIVES

There are 3 main incentive structures that can be used to encourage recruitment and participation: personal; social; and financial (Chamberlain, Poesio & Kruschwitz 2009). All incentives are applied with caution as rewards have been known to decrease annotation quality (Mrozinski, Whittaker & Furui 2008).

A previous survey showed that women are more likely to play social network games than men⁹ and this could have an impact on the types of incentives offered. Facebook allows access to user data, including gender, as part of the basic access for the application so gender and workload of the players could be investigated.¹⁴

Personal incentives

Personal incentives are evident when simply participating is enough of a reward for the user. Generally, the most important personal incentive is that the user feels they are contributing to a worthwhile project.

Also important for the players of *Phrase Detectives* is that they read texts that they find interesting. The choice of documents is important in getting users to participate in the game, to understand the tasks and to keep playing. Whilst some texts are straightforward, others can provide a serious challenge of reading comprehension and completion of linguistic tasks.

Texts were graded on complexity (on a scale of 1 to 4) on import. A player can choose the maximum level of document complexity they wish to read as they may be

¹⁴It was assumed for the purposes of this investigation that Facebook users declare their gender truthfully.

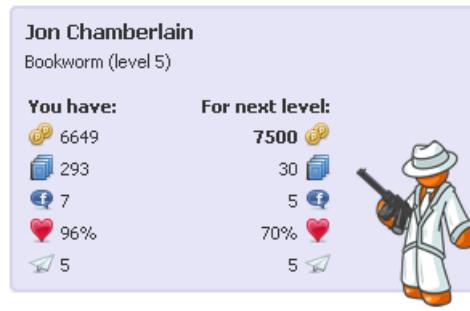


Figure 5. Detail showing criteria for the next level, displayed to the player on their homepage.

motivated to play the game to improve their English skills or, equally, because they enjoy reading challenging texts.

Social incentives

Social incentives reward users by improving their standing amongst their peers (in this case their fellow players and friends).

Using leaderboards and assigning levels for points can be an effective motivator, with players often using these as targets i.e., they keep playing to reach a level or high score before stopping (von Ahn & Dabbish 2008).

To investigate this in *Phrase Detectives*, players were grouped by how much progress they had made towards the next level in terms of points they had scored beyond the requirement for their current level. For example if a player had 110 points on level 2 (where the points requirement is 100) and the level 3 requirement is 200 points, then this player has progressed 10% towards the next level.

News feed (or wall) posting is integrated into the game. This allows a player to make an automatically generated post to their news feed which will be seen by all of the player's friends (see Figure 6).¹⁵

The posts include a link back to the game. Players are required to make a post from the game every time they are promoted to the next level. Posting is a very important factor in recruiting more players as studies have shown that the majority of social game players start to play because of a friend recommendation.^{9 16}

Posts may be social (display information about the document the player is working on or has just completed), collaborative (asking friends to join the game) or competitive (position in a leaderboard). Social posts are similar to information social network users share with friends so it is reasonable to assume they will be the most common type of post made from the game. This

¹⁵Since the release of the game Facebook has changed how posts are displayed. Posts from the game now appear on the player's profile and in a news ticker.

¹⁶<http://www.lightspeedresearch.com/press-releases/it's-game-on-for-facebook-users>



Figure 6. Detail of a news (or wall) post created automatically from the game, as seen by the player's friend.

was investigated by analysing the logs of wall posts made from the game.

Financial incentives

Financial incentives reward effort with money. When financial rewards were offered in the game they were for monthly high scores, ranging from £50 for first place, £15 for second place, £10 for third place and £5 for fourth to tenth place. During July 2011 additional daily lottery-style prizes of £5 were awarded, along with £20 for the player with the highest level, highest rating and largest team. The monthly prizes motivate the high-scoring players to compete with each other by doing more work and motivate low-scoring players in the early parts of the month when the high score is low. The prizes were sent as Amazon vouchers by email to the winners.

Whilst financial incentives are important to recruit new players, a combination of all 3 types of incentives is essential for the long term success of a game (Smadja 2009).

The effectiveness of incentives was analysed by looking at new players, active players and new annotations each month. Months where prize funds were available were compared to months where there was none and a per-annotation cost effectiveness was calculated. For the first 5 months no prize funds were offered but the following months all had prize funds of £110, except July 2011 which had £320 (see Table 1).

The site was promoted in February, July and December 2011 to email lists, existing players, relevant Facebook groups etc, as well as advertised on GoogleAds and Facebook (a modest total pay-per-click budget of £160) and competition websites.

RESULTS

Results from the Facebook interface of *Phrase Detectives* were analysed from February 2011 to February 2012.

Gender and workload of players

The Facebook game attracted 612 players of which 63% were female, 35% were male and 2% did not disclose their gender. Of the ten highest scoring players, 60% were female, 30% were male and 10% did not disclose their gender. This supports the previously mentioned survey that social network games are played predominately by women.

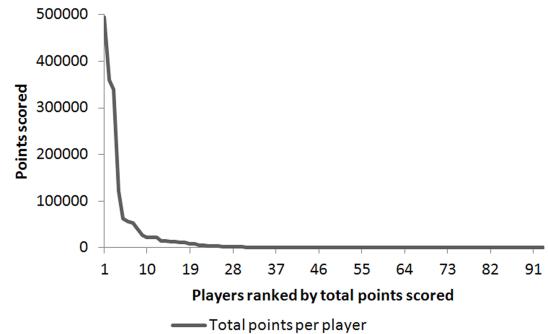


Figure 7. Chart showing the scores of players.

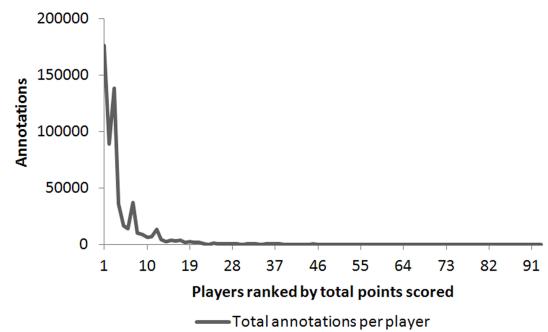


Figure 8. Chart showing the total annotations of players, ranked by their total score.

In a study of the previous version of *Phrase Detectives* it was reported that the ten highest scoring players (representing 1.3% of total players) had 60% of the total points on the system and had made 73% of the annotations¹⁷ (Chamberlain et al. 2009). In the Facebook version of the game the ten highest scoring players (representing 1.6% of total players) had 89% of the total points and had made 89% of the annotations (see Figure 7). The same ranking was used in Figure 8 to show that although the total number of annotations of players is generally in line with points scored it is not always the case.

These results show that the majority of the workload (in this case annotation of documents) is being done by a handful of players. However, the influence of players who only contribute a little should not be undervalued as in some systems it can be as high as 30% of the workload (Kanefsky, Barlow & Gulick 2001) and this is what makes the collective decision making robust.

A subset of 112 players (those that declared their gender, had a rating above zero and had completed at least one annotation) were analysed to investigate whether gender was related to the amount of work a player does. Of these players men represented 35% of the total (39 players). On average male players had completed 1,290 annotations and scored 4,636 points, compared to female

¹⁷For the purpose of data analysis, annotations and validations are counted together and referred to as annotations.

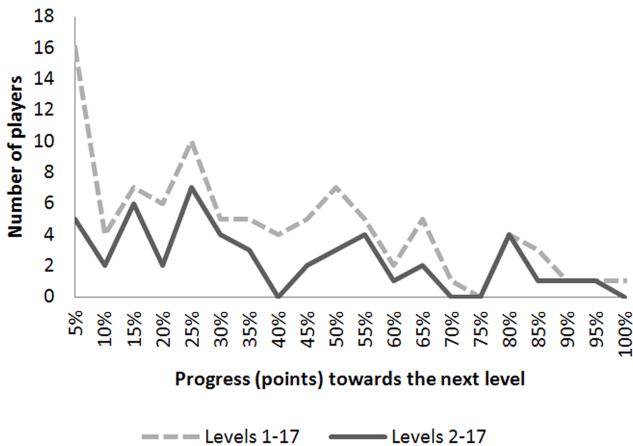


Figure 9. Chart showing the distribution of players' progress towards the next level.

players who had completed 4,636 annotations and scored 20,776 points. Although female players appear to be doing more work on average the difference is not statistically significant (using an unpaired t-test for significance testing).

The ten highest scoring male players were then compared to the ten highest scoring female players. On average, the former had completed 4,817 annotations and scored 17,628 points, whereas the latter had made 48,359 annotations and scored 144,905 points. The difference in both workload and score is statistically significant ($p < 0.05$).

This suggests that not only are female players more likely to play a socially networked game, they are also more likely to actively participate than male players. Further analysis is required to investigate whether female players provide higher quality annotations.

Using levels as goals

All players who had a score (i.e. they had passed the training stage and had completed a document) were categorised in terms of their progress to the next level. In the first analysis all players from levels 1 to 17 were included.

To distinguish between players who gave up after completing the training (i.e., the early stage of level 1) from the players stopping after reaching their goal of the next level, a second analysis was made excluding players on level 1.

The clustering of players near to the early stages of progress to the next level (see Figure 9) could be an indication that players are motivated to work towards the next level before stopping, however the level criteria and scoring system make this difficult to assess. Players score points in batches because all points scored on a document are added when the case is completed. Also, at this time, all extra points from other players who have agreed with their annotations are added. This

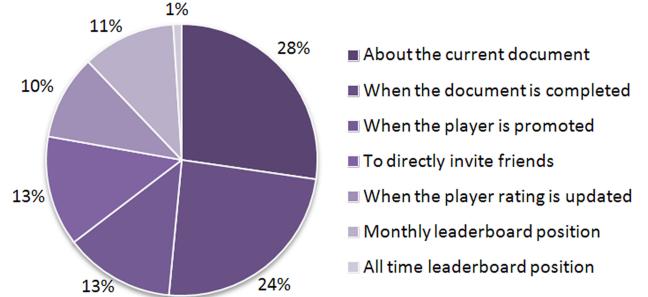


Figure 10. Chart showing the breakdown of 423 wall posts made by players.

effect becomes more negligible at higher levels where the progression to the next level is longer.

The analysis looks at player progression at the end of the data collection period i.e., players who get to the next level then never play the game again. To investigate this in detail the progression of each player session should be plotted however the game is not designed to test this and, if it were, many more players would be needed.

Whilst it is intuitive to think that players will use level thresholds as goals this evidence does not support it.

Posting from the game

Players' posts (see Figure 10) were most commonly social i.e., about the document the player was working on or had just completed (52%). This compares to competitive posts when they went up a level (13%), when their rating was updated (10%) or to post about their position in the leaderboard (12%). The remaining 13% of news posts were players making a direct collaborative request for their friends to join the game.

These results support the assumption that players are most likely to make posts from the game when the information is similar to what they might usually post. This should be a design consideration when developing a CI system for social networks.

The effect of incentives on recruitment and participation

Figure 11 shows the growth of *Phrase Detectives* on Facebook. Months where there was active promotion of the site (February, July and December 2011) show increases in new players, as one would expect. The month with the largest prize fund also had the most recruitment, double that of a later month with a smaller prize fund.

Based on the assumption that the first promotion month, when the site went live, was an exception as players of the previous game joined the new version, there is an indication that financial incentives increase recruitment to the game if sufficiently advertised.

It is noticeable that the amount of active players (a player who made more than one annotation in a particular month) stayed consistent and does not seem to increase

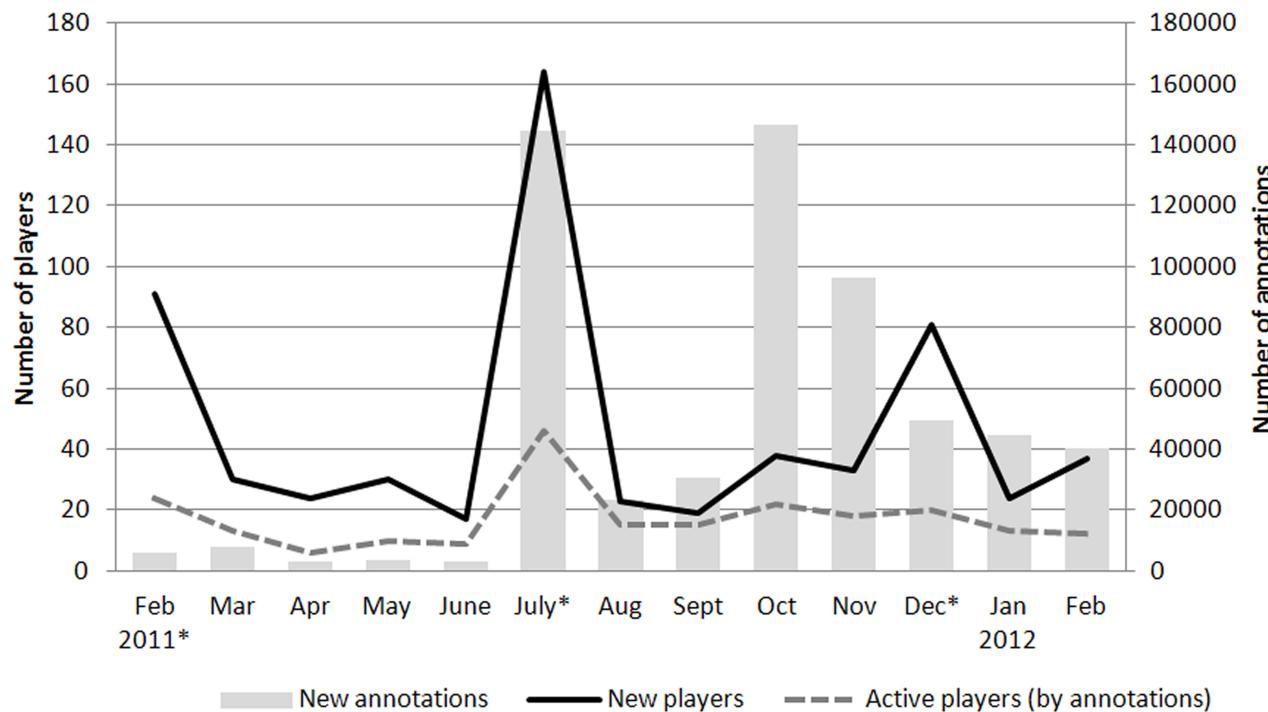


Figure 11. Chart showing new annotations plotted against new players and active players. Prizes were available in the game from July 2011 to February 2012 (see Table 1). * indicates a month with active promotion for the game.

with recruitment or financial incentives. Whilst it could be expected that the number of active players steadily increases over time as more players are recruited, the results show that most players will play the game for a short period of time and only small number continue to play every month. This is confirmed by the previous results that only a handful of players do most of the work.

Financial incentives do appear to be a strong motivating factor when considering how much work the active players do. Months with prizes have considerably more new annotations than those without, but with a similar number of active players.

This suggests that active players are motivated by financial incentives, however the large amount of gameplay in October and November 2011 indicate that other motivating factors such as personal and social incentives are, to some extent, also successful.

Cost effectiveness of financial incentives

The cost per annotation is a measure of the effectiveness of the prize fund. A baseline of 4699 new annotations per month could be expected without a prize fund (the average of the first 5 months' new annotations) so anything over that could be attributed to the prize fund.

The average cost per annotation across the months where there was a prize fund was £0.0028 (see Table 1) and this

can be compared to other CI systems. Amazon Mechanical Turk (MTurk)¹⁸ is often used as a way of getting data quickly but can cost £0.007 - 0.70 (\$0.01 - 0.10) per annotation (Mason & Watts 2010).

Both systems require the annotations to be aggregated before a collectively intelligent answer can be produced however even professional annotation schemes require some degree of validation.

The advantage of GWAP over MTurk is that personal and social incentives can be used, as well as financial, to minimise the cost and maximise the persistence of the system.

It is also worth considering the setup and maintenance costs of CI systems in a cost per annotation analysis.

CONCLUSIONS

In addition to accessing the vast human resources that social networking platforms offer, CI systems can also take advantage of the inherent social structure and shared personal data available to maximise the incentives that encourage participation.

Phrase Detectives has shown some valuable insights into user motivations in social network gaming and participation in CI efforts. The results support previous surveys that show women are more likely to play, and will spend more time playing, socially networked games.

¹⁸<http://www.mturk.com>

Table 1. Cost per annotation based on the prize fund.

Month	Prize fund	Cost/annotation
Feb - Jun 2011	-	-
Jul 2011	320	0.0022
Aug 2011	110	0.0058
Sep 2011	110	0.0042
Oct 2011	110	0.0008
Nov 2011	110	0.0012
Dec 2011	110	0.0025
Jan 2012	110	0.0028
Feb 2012	110	0.0031

There are indications that players want to share information from the game that is similar to the information usually shared in their social networks. The success of a socially networked game relies on creating an experience that players want to share with their friends.

The results suggest that attracting and motivating the right kind of player is as important as attracting lots of players because, although collective intelligence needs a crowd, that crowd also needs to do some work. Financial incentives, coupled with promotion, increase recruitment and have a considerable impact on participation.

The increase in data collected from the game due to financial incentives still makes it a cost effective alternative to other CI systems such as MTurk.

ACKNOWLEDGMENTS

Thanks to Ans Alghamdi for his work on the Facebook version of the game. The creation of the original game was funded by EPSRC project AnaWiki, EP/F00575X/1.

REFERENCES

- Albakour, M.-D., Kruschwitz, U. & Lucas, S. (2010), Sentence-level attachment prediction, in ‘Proceedings of the 1st Information Retrieval Facility Conference’, Vol. 6107 of *Lecture Notes in Computer Science*, Springer, Vienna, pp. 6–19.
- Chamberlain, J., Poesio, M. & Kruschwitz, U. (2008), Phrase Detectives: A Web-based Collaborative Annotation Game, in ‘Proceedings of the International Conference on Semantic Systems (I-Semantics’08)’, Graz.
- Chamberlain, J., Poesio, M. & Kruschwitz, U. (2009), A new life for a dead parrot: Incentive structures in the Phrase Detectives game, in ‘Proceedings of the WWW 2009 Workshop on Web Incentives (WEBCENTIVES’09)’, Madrid.
- Feng, D., Besana, S. & Zajac, R. (2009), Acquiring high quality non-expert knowledge from on-demand workforce, in ‘Proceedings of the 2009 Workshop on The People’s Web Meets NLP: Collaboratively Constructed Semantic Resources’, People’s Web ’09, Association for Computational Linguistics, Morristown, NJ, USA, pp. 51–56.
- Johnson, N. L., Rasmussen, S., Joslyn, C., Rocha, L., Smith, S. & Kantor, M. (1998), Symbiotic Intelligence: Self-Organizing Knowledge on Distributed Networks Driven by Human Interaction, in ‘Proceedings of the Sixth International Conference on Artificial Life’, MIT Press.
- Kanefsky, B., Barlow, N. & Gulick, V. (2001), ‘Can distributed volunteers accomplish massive data analysis tasks?’, *Lunar and Planetary Science* **XXXII**.
- Mason, W. & Watts, D. (2010), ‘Financial incentives and the performance of crowds’, *ACM SIGKDD Explorations Newsletter* **11**(2), 100–108.
- Mrozinski, J., Whittaker, E. & Furui, S. (2008), Collecting a why-question corpus for development and evaluation of an automatic QA-system, in ‘Proceedings of ACL-08: HLT’, Association for Computational Linguistics, Columbus, Ohio, pp. 443–451.
- Rafelsberger, W. & Scharl, A. (2009), Games with a purpose for social networking platforms, in ‘Proceedings of the 20th ACM conference on Hypertext and hypermedia’, ACM, pp. 193–198.
- Singh, P. (2002), The public acquisition of commonsense knowledge, in ‘Proceedings of the AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access’, Palo Alto, CA.
- Smadja, F. (2009), ‘Mixing financial , social and fun incentives for social voting’, *World Wide Web Internet And Web Information Systems* (April).
- Snow, R., O’Connor, B., Jurafsky, D. & Ng, A. Y. (2008), Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks, in ‘EMNLP ’08: Proceedings of the Conference on Empirical Methods in Natural Language Processing’, Association for Computational Linguistics, Morristown, NJ, USA, pp. 254–263.
- Surowiecki, J. (2005), *The Wisdom of Crowds*, Anchor.
- von Ahn, L. (2006), ‘Games with a purpose’, *Computer* **39**(6), 92–94.
- von Ahn, L. & Dabbish, L. (2008), ‘Designing games with a purpose’, *Communications of the ACM* **51**(8), 58–67.