Snag'em: Graph Data Mining for a Social Networking Game

Veronica Cateté
North Carolina State University
911 Oval Drive
Raleigh, NC 27606
vmcatete@ncsu.edu

Drew Hicks
North Carolina State University
911 Oval Drive
Raleigh, NC 27606
aghicks3@ncsu.edu

Collin Lynch
North Carolina State University
911 Oval Drive
Raleigh, NC 27606
cflynch@ncsu.edu

Tiffany Barnes
North Carolina State University
911 Oval Drive
Raleigh, NC 27606
tmbarnes@ncsu.edu

ABSTRACT

New conference attendees often lack existing social networks and thus face difficulties in identifying relevant collaborators or in making appropriate connections. As a consequence they often feel disconnected from the research community and do not derive the desired benefits from the conferences that they attend. In this paper we discuss Snag'em, a social network game designed to support new conference attendees in forming social connections and in developing an appropriate research network. Snag'em has been used at seven professional conferences and in four student settings and is the subject of active research and development. The developers have sought to make the system engaging and competitive while preventing players from 'gaming' it and thus accruing points while neglecting to form real-world connections. We briefly describe the system itself, discuss its impact on users, and describe our ongoing work on the identification of critical *hub* players and important social networks.

Keywords

Social Networks, Gamification, Conferences, Underrepresented Populations

1. INTRODUCTION

Social networking is an essential task at any academic conference or professional venue. One of the primary goals of attendees is to seek out relevant work, identify potential collaborators, and to maintain existing connections. Many of these contacts are made by building upon existing relationships and by expanding the attendees existing social network. New conference goers however, particularly students and historically underrepresented groups, lack these foundational networks and thus face difficulties making connections. Based on Tinto's Theory of University Departure, increased interaction with other students, faculty, staff and community supporters can increase the retention rate of minority populations and sense of community within secondary and post-secondary academic communities [7].

In academia, sense of community has a strong positive correlation with retention [7]. Research indicates that students who do not feel as if they are part of a larger academic community are less likely to participate in extracurricular activities and organizations. This leads to lower retention rates, especially amongst minority students who suffer without a strong student support group [7]. A feeling of community can be nurtured with small group activities that augment the individual's role within a setting and helps students to foster connections [8].

Snag'em was designed as a pervasive game to encourage valuable professional networking and promote sense of community. The system's pervasive features are designed to help players translate their in-game networks directly into real world peer groups. The system was originally created for the 2009 Students and Technology Academia Research & Service (STARS) conference. This conference is unusual in that it is an academic conference designed specifically to engage with minority and female undergraduates majoring in computing fields. Students who attend the conference participate in competitions and attend training sessions to support engagement and research. Studies conducted at prior conferences has shown that while students were engaged in the training sessions and vigorously involved in learning they did not develop the lasting social connections that can arise out of conferences. Snag'em was designed to engage students in social networking through gamification of the process. Prior research has shown that social games can help people to engage in otherwise challenging or uncomfortable situations [6, 4, 2, 3].

Snag'em functions as a large human scavenger hunt. Players are assigned a set of relevant tags (e.g. "I'm a games researcher", or "I'm interested in data-mining"). They are

Figure 1: The browser interface for mission assignments. Snag Snapshots highlight missions recently completed.

then assigned a set of missions (e.g. "Find someone who specializes in HCI") which they must complete by identifying and engaging with an appropriate individual. The system was developed in PHP with a MySQL backed and provides a web-based front end for players to edit their profile and to record interactions. We have also developed a mobile version of Snag'em which allows players to access the game via tablets and smartphones. The game itself is designed for easy deployment to new conferences and we are presently adding features that will allow us to automatically populate the database with initial tags.

Figure 1 shows a snapshot of the mission browser screen from the web version of Snag'em. Contact is registered when the players enter a 4-digit ID from the other person. In addition to missions the systems also allows players to record notes about one-another for future reference (e.g. "I should e-mail my proposal to him after the conference") and to send one-another messages. A sample message from the mobile interface is shown in Figure 2. Snag'em can also be configured to suggest specific individuals that students should make contact with based upon their mutual interests or social connections.

The system logs all player interactions including tag updates, missions completed, notes made, messages, sent, connections added, and so on. This provides a rich dataset of information that we can use to analyze social patterns at conferences and to improve the impact of the intervention. In addition to the raw logs the game contains a number of features to support easy analysis. The developers have created a set of badges that allowed administrators to easily track the number of people playing via the mobile or web interfaces as well as the number of missions completed. The badge system also provides a simple visual record of the types of features (i.e. notes, tags, avatars) each player is using. The badge systems also allows administrators to note the frequency of use, time of day that players are online and so on.

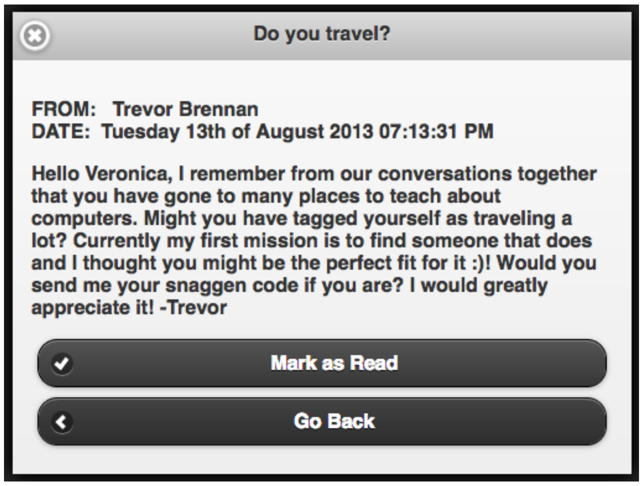

Figure 2: Here is an example of a message sent in game after a conversation between players.

To date, Snag'em has been used at seven academic conferences. It has also been deployed to help incoming freshman and transfer students connect at four academic institutions. In 2009, for example, Snag'em was used by new students in the College of Computing and Informatics at the University of North Carolina at Charlotte. Students were able to play the game during the freshman orientation week with kiosks available for students to sign up located in the College of Computing and Informatics. SNAG'EM was used alongside other social activities to get students acquainted with each other, the faculty, and the CCI campus.

2. PRIOR ANALYSIS

We have studied the impact of Snag'em on users and found that playing the game improved conference attendees' sense of community [6, 1]. We have also analyzed the existing dataset both to test the implementation of the Snag'em features, and to identify *hubs* or critical players whose activity predicts the behavior of others.

In analyzing the game mechanisms we have focused primarily on the STARS 2009 dataset. As mentioned above STARS is primarily targeted at undergraduate students specifically females and underrepresented minorities. We deployed the system via the conference infrastructure and set up a table near the registration booth. The game was active during the first two full days of the conference. The conference had 280 attendees 60.0% of whom were female (N=168) and 70% of which (N=196) were students. Roughly 28% of the conference-goers played the game (N=80) of whom 50% were female. In previous analysis 35.0% of the players were classified as active. It is important to note that this data was collected on an earlier version of SNAG'EM where players could snag each other only once, and only a single mission was available at a time. Because completing missions was significantly more difficult in this version of the game, players were classified as active if they completed at least two missions. An additional 50% of the players were classified as Interested, meaning they did more than just register for the game or that they completed one mission.

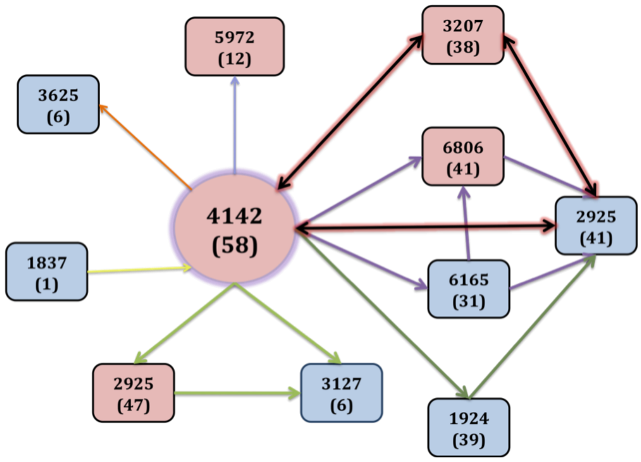

Figure 3: Visualization of community center 4142, with one of that user's maximal cliques highlighted.

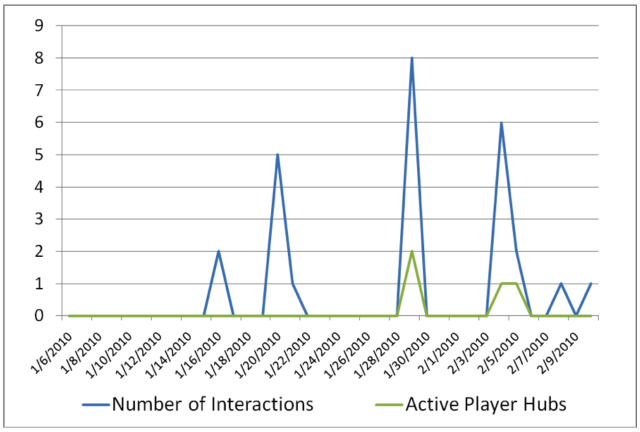

Figure 4: Correlation between active player hubs and number of interactions.

Our analysis of this data was focused primarily on the mission and scoring systems. In 2009 the mission system was relatively simple and focused solely on guiding students to locate a single individual with a desired tag. Players were then guided to record the match via the ID system discussed above. Both the missions generated and points received were determined by the state of the current network. When generating missions we attempted to ensure that they were of varying difficulty, and were relevant to the current user. In this iteration of the system the missions could only be satisfied by identifying someone whom the user had not previously snagged. The target tags were selected from the full set listed in the system. Easy missions were assigned high frequency tags (more than $\frac{1}{2}$ of the non-adjacent users), while medium missions were assigned tags that are present in $\frac{1}{4}$ of non-adjacent users and hard missions required tags present in less than $\frac{1}{4}$ of the non-adjacent community.

The difficulty of the mission determined the base score which was then modified by a *connectedness factor*. This factor was greater than 1 if adding this connection expanded your "Friends of friends," that is, the number of vertices less than 2 edges distant from the user. The connectedness factor was less than 1 if you completed the mission using the ID of a person you were already adjacent to, In this way we hoped to encourage players to branch out.

When developing the system we had hoped that players would develop social networks that exhibited breadth (i.e. meeting lots of people), depth (i.e. getting to know some individuals well), and mutuality (i.e. snags in both directions). We therefore hoped that users' immediate neighborhoods would be large and relatively dense with multiple snags between some people and bidirectional connections. When analyzing the STARS 2009 dataset, however, we found that this was not the case. Rather the game mechanics encouraged players to make a relatively large number of unrelated connections which, in turn, produced relatively broad and shallow social neighborhoods with very few inbound arcs. In fact some players actually opted to hide their IDs so that no other player could gain points by using them to complete a mission. As a consequence the attendees were actually less likely to engage in the deep and meaningful conversations required or to form lasting connections.

In response to these results we have overhauled the scoring system. This included changing the connectivity bonus to reward players based upon the size of the largest clique that they participate in. Players are now rewarded more for expanding this clique, thus deepening their social networks, than they are for adding an unrelated individual to their friends of friends. We have also allowed players to re-snag the same individual for multiple missions with a low penalty for re-snags, and have begun to reward players with points for allowing themselves to be snagged to help others complete a mission. We have not yet analyzed the effects of these changes on a the dataset.

We have used two measures of importance when identifying critical players. The first is the simple interaction frequency as measured by the number of outgoing arcs from a player in the network. The second is membership in maximal cliques, that is, cliques which are not part of a larger clique. Players that participate in a large number of maximal cliques are *hubs*. We were able to identify three distinct user communities in the STARS 2009 dataset that centered on these hubs. A sample community graph is shown in Figure 3. We also found that the activity of these hub players was highly correlated with the activity of the other players in the community (r=0.827). A graph of these spikes is shown in Figure 4. More specifically, on any day where one or more of the hub players were active, we observed spikes in the number of interactions taking place across users. We were able to observe a similar effect (r = 0.659) on days when the developers had a booth/kiosk available.

We also performed an analysis of hub players using the UNCC Student Orientation dataset described above. In this dataset 91 of the 1290 potential students registered to play Snag'em of which 22% (N=20) were female [5]. This data was collected on a version of Snag'em permitting multiple missions and allowing players to connect with the same user multiple times.We classified players as active if they completed 5 or more missions. In total, 9 users were active users during this study. However, all of these players were moderators or members of the development team. In this deployment almost all of the game interaction took place at the registration table thus making the administrators responsible for most of the activity. We had hypothesized that the moderators would only need to initiate the game and then it would be self-sustaining. As our analysis shows however, this was not the case. In general the players did not think about the game outside of the advertised area.

3. OPEN QUESTIONS & FUTURE WORK

Our prior research has focused on identifying key players using graph methods. We plan to continue examining these key players in future work and to modify the mission selection criteria to better engage players that have not been active recently. Our chosen method of community detection, based upon maximal cliques, is both computationally expensive on large networks and can change substantially based upon small shifts in the network. Using a simpler, less volatile measure to identify community centers would allow us to adapt the gameplay based upon those communities more efficiently. This would in turn enable us to encourage new players to specifically seek out these active players in an effort to better engage them from the start. Different community detection algorithms might identify different hub players, or provide different ways of scoring missions that help to foster larger communities. Further development in this area might facilitate play in the absence of an instigating 'active player' or outside of areas with an active game station or kiosk.

One open question is how to better identify hub players during the game, and modify mission selection criteria to engage inactive players or players who don't need motivation to network. These 'social elites' are important to attract, as they are precisely who we should be encouraging our players to network with. If we are better able to build and analyze our networks, we may be able to offer features to these social elites that would attract them to Snag'em as a system more than the gamification aspects would. We hope to explore techniques for reliably generating edges and tags for users based on existing data sources like conference proceedings or citations. This would reduce the burden of entry on new players, particularly elites, and make it more likely for those users to participate in networking (if not gameplay) using SNAG'EM.

We also plan to expand our in-game evaluation of Snag'em itself. We are presently adapting the system to poll players for their opinions as the system is used. This will better help us to identify the immediate impact of the system on users' social connections. We will be deploying some of these new features of the system during the 2014 Educational Datamining Conference in London as well as subsequent conferences in 2014 and 2014.

4. ACKNOWLEDGMENTS

This research was supported by the NSF GRFP Fellowships No. 0900860 & No. 1252376 and BPC Grant No. 0739216 and No. 1042468 Thanks to all developers who have worked on the SNAG'EM project. The authors also wish to thank Shaghayegh Sahebi for her expert advice.

5. REFERENCES

[1] S. L. Finkelstein, E. Powell, A. Hicks, K. Doran, S. R. Charugulla, and T. Barnes. Snag: using social networking games to increase student retention in computer science. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*, pages 142–146. ACM, 2010.

[2] M. Montola. Exploring the edge of the magic circle: Defining pervasive games. In *Proceedings of DAC*, page 103, 2005.

[3] M. Montola. A ludological view on the pervasive mixed-reality game research paradigm. *Personal and Ubiquitous Computing*, 15(1):3–12, 2011.

[4] E. Powell and T. Adviser-Barnes. A framework for the design and analysis of socially pervasive games. 2012.

[5] E. Powell, F. Stukes, T. Barnes, and H. R. Lipford. Snag'em: Creating community connections through games. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 591–594. IEEE, 2011.

[6] E. M. Powell, S. Finkelstein, A. Hicks, T. Phifer, S. Charugulla, C. Thornton, T. Barnes, and T. Dahlberg. Snag: social networking games to facilitate interaction. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4249–4254. ACM, 2010.

[7] V. Tinto. Taking Student Retention Seriously: Rethinking the First Year of College. *NACADA Journal*, 19(2):5–10, 2000.

[8] S. White. Algorithms for estimating relative importance in networks. *Proceedings of the ninth ACM SIGKDD international*, pages 266–275, 2003.