Going Out of Business: Auction House Behavior in the Massively Multi-Player Online Game *Glitch*

Abstract

The in-game economies of massively multi-player online games (MMOGs) are complex systems that have to be carefully designed and managed. This paper presents the results of an analysis of auction house data from the MMOG *Glitch*, across a 14 month time period. The data comprise almost 3 million data points, over 20,000 unique players and more than 650 products. Furthermore, an interactive visualization, based on Sankey flow diagrams, is presented which shows the proportion of the different clusters across each time bin, as well as the flow of players between clusters. The diagram allows evaluation of migration of players between clusters as a function of time, as well as churn analysis. The presented work provides a template analysis and visualization model for progression-based or temporal-based analysis of player behavior broadly applicable to games.

1. Introduction

MMOG economies are complex systems that have to be carefully designed and managed to ensure that they properly support the population and provide useful functionality for generation, exchange and elimination of resources (Bartle 2003; Castronova et al. 2009; Bloomfield and Young-Jun 2011; Plumer 2012). In conjunction with the social and combat mechanics of an MMOG, these systems form the backbone of these games. From the perspective of game development, MMOG economics are vital and has in recent years seen professional economists getting employed with MMOG (and other game-) companies to design and manage the complex and dynamic economies of virtual worlds (see e.g. Plumer 2012; Shokrizade 2013; Grabianowski 2013).

The work presented here is situated within the domain of economics-focused game analytics, focusing on combining dimensionality reduction of economic data with temporal pattern identification. There are two overall goals: 1) to provide a longitudinal analysis of the auction house data from the Free-to-Play, browser-based MMOG *Glitch* with the purpose of uncovering any patterns in the player behavior in connection with the in-game economy of the game. The analysis combines dimensionality reduction techniques with temporal analysis to enable the definition of clusters of player behavior, and how these evolve across time. 2) To develop a novel, interactive visualization of the data, which enables non-experts to interact with data and results and draw their own conclusions.

The motivation for the first goal is founded in previous work on online economies but departs from the majority of previous work in that it does not attempt to fit e.g. a financial theory to a MMOG market or test specific hypotheses, but rather works in a data-driven, explorative fashion. The motivation for the second goal is founded in the increasing need in the industry (and game science) for techniques that allow analysts to provide analysis results to a variety of stakeholders in the industry – including game designers, producers, marketing, community management etc. – in a way that allows these varied groups to dynamically interact with the data, without requiring extensive training or data mining expertise (Seif El-Nasr et al. 2013). Data visualization is a well-established field in its own right (Tufte 1983), but the history of applying theories and techniques from data visualization to the specific context of game analytics is more recent (Medler 2012).

2. Contribution and results

The work presented consists of two parts: 1) A cluster analysis of how *Glitch* players utilized – with varying degrees of success – the auction house, across a 14 month time period. The data comprise almost 3 million data points, over 20,000 unique players and more than 650 products. For the cluster analysis, 5 Key Performance Indicators (KPIs) were defined to form as the basis for the analysis. K-means clustering was applied across 14 time bins, enabling analysis of the temporal pattern of how the different clusters performed.

2) An interactive Sankey diagram, which shows the proportion of the different clusters across each time bin, as well as the flow of players between clusters. The diagram allows evaluation of migration of players between clusters as a function of time, as well as churn analysis. The diagram is available on: http://powerful-meadow-8588.herokuapp.com/.

In combination, the analysis and visualization combines the considerations of behavioral economic analysis with data visualization, and provides a template model for conducting progression/temporal-based analysis of player behavior in games - across any dimension of behavior.

3. Previous work

Given the space constraints imposed here, this section will focus on the research most relevant to the work presented here: From the perspective of the industry, the systems and mechanics established for controlling the flow of resources within a complex game environment are foundational to the success of a game, and therefore the investigation of virtual world economics essential to the game industry (Bartle 2003; Castronova et al. 2009).

From the perspective of academic research, in-game economies are of interest perhaps most importantly because MMOGs and other forms of massively multi-user virtual environments form semi-controlled environments that allow investigation of human behavior – and associated economic modeling – at a societal scale. Previously, behavioral economics has been largely limited to building simulations (Plumer 2012). This also holds in a games context due to a lack of access to controlled environments or large-scale datasets (Drachen et al. 2009; Keegan et al. 2010). This is also the case with in-game auction houses, which unlike real-world auction sites such as eBay; do not contain any information asymmetries between buyers and sellers. Though there are vast differences between different kinds of items in *Glitch,* all items of the same kind are completely identical. Furthermore, information about items is freely available (viewable by "mousing over" the item or selecting it). The ability of a MMOG developer to also track and analyze behavioral data and demographic data about the player population, has opened up new ways for economists to study consumer behavior (e.g. Castronova et al. 2009; Seif El-Nasr et al. 2013). Several authors have examined auction house data from MMOGs specifically, including Simpson (1999), whose early work on *Ultima Online* (one of the first MMOGs) described an environment with hundreds of different items being exchanged via virtual auction houses, and highlighted the macroeconomic design built into the game from the onset to facilitate economic flow. Importantly, Simpson (1999) highlighted the failure of a shopkeeper-run (not player controlled) economy, which led to *Ultima Online* experimenting with player-run auctions, i.e. more open economic systems, forming the foundation for this feature in future games. More recently, Castronova et al. (2009) investigated macroeconomic behavior in *EverQuest II* working with transaction data, proposing an empirical test of whether aggregate economic behavior maps from the real to the virtual world. The authors noted that an aspect of virtual world auction data that makes adoption of real-world aggregates in the virtual domain easier, is that items listed for sale in *EverQuest II* are simple to map to real-world analogs, e.g. food items, furniture items, etc. This is also the case for e.g. *World of Warcraft* and *Glitch*. Morrison and Fontenla (2013) studied eight *World of Warcraft* servers, and noted that price convergence occurred at all of these. Finally, Lohdonvirta (2009) used virtual item sales records to identify attributes that drive purchase decisions.

4. Glitch

Glitch was a browser based MMOG developed by Tiny Speck. The majority of the games lifetime was in beta, with a short stint as public release from September 27, 2011 to November 30, 2011. Tiny Speck announced on November 14, 2012 that the game would close December 9, 2012. At the time of public launch the game had 27,000 active users and a free to play model with no transactional elements in place.

Glitch's core gameplay revolved around crafting, resource gathering, trading and social elements in an open-ended world. The main objective of the game was to help build the world and create mini-games within. Structural elements like leveling and questing existed to facilitate the acquisition of resources needed to fully participate in the world-building mechanics. The in-game currency, Currants, were obtained through questing, various locations throughout the world, selling items to NPCs, and selling items to other players privately or via the auction house. Similar to other in-game MMO auctions, players could post any quantity of an item on the market and see other player's posting of the same items: *"Postings expired after three days, and Tiny Speck would claim a small fee for each posting. Initially listing an auction cost a 3% Currant fee, followed by an 8% commission fee (in Currants). On May 25, 2012 commissions on auctions were dropped and listing fees were adjusted upward to 7%"* (Landwehr 2013). Third party developers also created tools for users to track items in the auctions house when away from the game via smartphones and tablets.

5. Data

From shortly following launch until the game shut down, Landwehr (2013) collected telemetry data from the game. The data collected focus on auction house behavior, notably sales between players and records of item prices in the game, but also other data, e.g. on friendships, which is however not used here due to lack of temporal information.

During the relatively brief life of *Glitch*, thousands of players engaged in financial transactions via the games native auction house. The data contain an account of the behavior of the players in relation to trade that cross the lifetime of a MMOG. Such a dataset have not previously been available for study outside of the companies that run and manage MMOGs. The collected data centered on in-game auction and economic activity, as well as general forum discussions and friend networks. Using the game's built-in API for 3rd party developers, Landwehr (2013) scraped the data using a mix of Python scripts and Curl HTML scraping tools to mine four key areas of data: auction sales data, item street prices, forum conversations, and in-game friendship networks. Data is publically available in CSV, HTML XML formats.

The auction sales data was retrieved via a Python script that pulled the list of active auctions one every 10 seconds into a MySQL database from November 17, 2011 to December 10, 2012. Landwehr (2013) noted that several intervals were lost due to disk space/memory issues. Thus the data set cannot be considered completely comprehensive of every in-game auction posting. Approximately 3 million auctions were collected over 14 months.. For every auction posted the following data fields were collected: player id, timestamp, action expiration date, item name, item category, item quantity, tool uses, tool capacity, and the final outcome of the auction (sold, expired, or deleted).

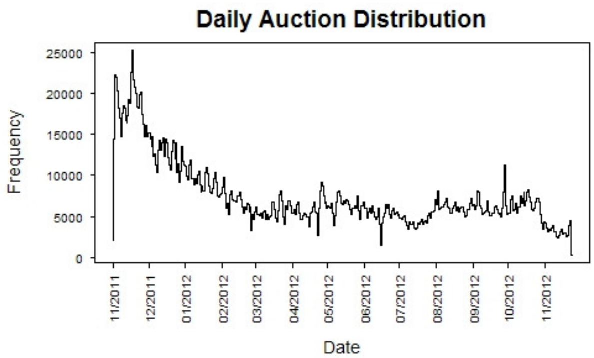

Figure 1: Distribution of daily auction activity

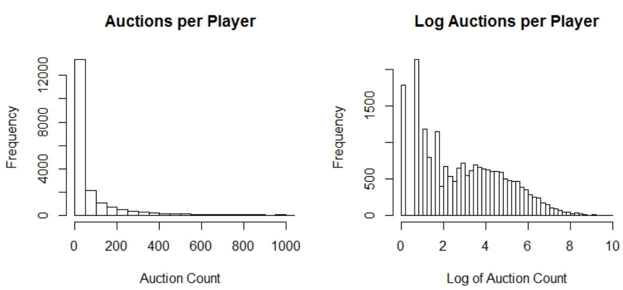

Figure 2: Distribution of auctions per player

The data scraped from the forum was provided in two forms: 1) a CSV file of player id, timestamp, and comment index number 2) separate HTML files corresponding with each comment index with the forum posting text.

For our analysis, we focused on the in-game auction activity using auction sales data and forum posting related to the game market place, as a proxy to better understand overall game health in relations to player activity. There were 2,914,359 data points on the auctions table, where we see 20,266 unique players listing 679 unique products across 41 unique categories from 11/17/2011 to 12/10/2012. 80% of total auctions are in 10 (41 total) categories, with 62% of total auctions coming from the top 5 categories. More granular, 28% of total auctions are in 10 products, with "meat" being the most popular auction item - 8% of all auctions.

There was an average (μ) = 7,472 auctions per day with a standard deviation (σ) = 4,081. The high was seen on 12/04/2011 with 25,252 auctions created, while the low was during the final days of data collection on 12/10/2012 with 318 logged (Fig. 2). On average, there were 143 auctions listed per player participating in the auction house, with σ = 546. The range among the players was from 1 to 2,133. Finally, we saw 85% of auctions successfully result in a sale, with μ = 85% of successful auctions per month (SD 2.9% points) and μ = 35% success per player per month who posted (σ = 47.7%) (Fig. 2).

6. Methodology

The work presented comprised two components: cluster analysis and generation of the Sankey diagram. These are described separately.

6.1 Cluster analysis

To perform the cluster analysis 5 KPIs (key performance indicators) were defined that would be actionable for measuring the health of the in game economy: Total Auctions, Average Auctions Posted per Day, Percentage of Auctions that Resulted in Sales, Number of Distinct Items Categories Posted, Marketplace Forum Posting (Boolean). These calculated variables were aggregated from both the auctions data table and the forum posting data table. We chose each month as a natural point for partitioning the data. Across the 14 months of data collection, players participated in the auction house μ = 2.76 months, with σ = 2.5 months. 41% of the players participated in the auction system for only one month.

Cluster	μ cluster size across months	σ	Months manifested
Moderate	**0.16**	**0.00**	**2**
Forum	**0.04**	**0.01**	**13**
Hardcore	**0.24**	**0.02**	**14**
Casual Losers	**0.14**	**0.02**	**12**
Casual Winners	**0.22**	**0.02**	**12**
Moderate Farmers	**0.15**	**0.03**	**14**
Moderate Losers	**0.14**	**0.02**	**6**
Moderate Miscellanea	**0.16**	**0.02**	**9**
Causal Forum	**0.03**	**0.00**	**4**
Moderate Winners	**0.15**	**0.00**	**1**
Casual	**0.35**	**0.00**	**2**

Table 1: Cluster sizes and number of months manifested.

After testing for independence, we applied k-means clustering to each of the 14 months' worth of data. To compensate for right skewness in the data, input variables were standardized into quintiles. In determining the optimal number of clusters for each month, we set a standard for maximizing heterogeneity across clusters and homogeneity within clusters. We selected the size of the group based on a within cluster variance / between cluster variance ratio of 0.3. Using this method there were 5-7 clusters except the final two months of the game in which the total number of players dropped significantly.

While the number of active in-game auction house users diminishes over time, the relative size of each of the clusters remains relatively constant. See the below chart for cohort size over time. For example, over the 14 months of hardcore players represent 22% to 28% of the total population at any given month. Among those clusters that are manifest for the majority of the months, the Moderate Farmer cohort exhibits the greatest variance in cluster size across the months, while Forum posts are most consistent in their population size from month to month.

6.2 Player Type Descriptions

In full, four main player types emerged, Casual Players, Moderate Players, Hardcore Players, and Forum Players (Fig. 5). Within these larger types clear distinctions emerged, resulting in 11 distinct player types, discussed below. Those that remained at the most basic type (i.e. Casual, Moderate, Hardcore) occurred in months where there were too few players to create divergent clusters each month (these two months contained only 5 groups).

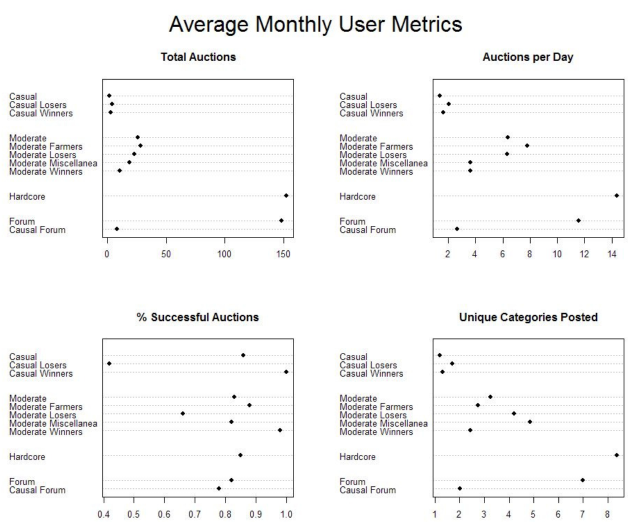

Figure 5: Average monthly values for each of the input variables in the cluster analysis grouped by cluster. While standardized values were used for the clustering the above chart represents the raw average for direct comparison.

Casual Players: Among casual players, two groups exist with key distinctions by auction success rate. Casual Losers have 42% sales rate on all auctions, while Casual Winners have 100% success rate (i.e. all auctions posted are sold).

Moderate Players: Moderate players are characterized by mid-range activity level. Moderate Farmers post a large number of auctions per day across a smaller range of item categories. These players are focused and establishing a trade niche within the economy. Moderate Miscellanea are less focused, instead posting auctions across a broader range of categories at a more leisurely pace (i.e. fewer auctions per day). Moderate Losers post more frequently than Moderate Miscellanea but are less successful than all other clusters in the moderate category.

Forum Players: Forum players have a similar economic play style to that of hardcore players with the exception of their posting activity in the marketplace forum. During the months of Feb – May 2012, a group of more casual players begin posting in the forums.

Hardcore Players: Hardcore players exhibit the least divergence in their group and across time. These players perform at the upper bound of all KPIs.

6.3 Sankey diagram design and development

Sankey diagrams comprise a series of nodes and links, similar to network diagrams. The key difference is that nodes are positioned or stacked at key intervals along one of the axes. For instance, at month t, players are categorized into 5-7 player behavior clusters. These players may have shifted clusters (or entered or left the game) from the previous period and into the next period. Nodes are positioned at months *1,2,…,14* while links show the shifts in players across clusters from one month to the next.

The diagram was developed using D3.js (Data-Driven Documents), a javascript library for binding data to HTML elements and Scalable Vector Graphics (SVG) shapes to dynamically render data visualizations (Bostock 2012).

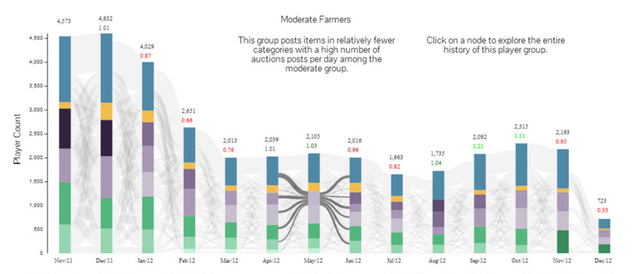

Figure 3: Glitch Sankey Diagram: Combining the click through elements of the Sankey diagram with standard bar chart formatting and measurement

Utilizing the Sankey diagram plugin for D3.js (http://bost.ocks.org/mike/Sankey/), we reduced padding, the amount of whitespace, between the vertically-placed nodes to replicate a stacked bar chart at any time *t*.

Hovering the mouse over any link or node shows the respective player count, cluster descriptions, and information about new and departing players. Total player count, a one-month retention rate, and x-axis labels were added via an in-file javascript variable. Additionally, a click-through of any node will provide a more detailed view of that group of players through the 14-month history.

The raw data to be plotted was a javascript object notation (JSON) file containing separate node and link information. A JSON file is a hierarchically structured document describing objects through the use of parents (objects) and children (descriptors). A given parent can contain multiple children and sub-children, and various data types can be exist in a single file. Our data contained two parents: nodes and links. The nodes parent contained children of month, cluster mapping, value, color, and departing and joining player counts, while the links children were source, target, and value. Links were mapped from source to target (with curvature defined in the Sankey.js file), and the stroke width was sized by the value of players moving from source to target. The nodes were sized based on the value noted in the JSON file, representing the size of players existing in that cluster in that month.

Without adjusting the Sankey.js file, which contains all the code for computing the node depths, link connections, etc., there are a few easy to adjust arguments when calling the Sankey() function: node width, node padding, and layout. To reduce the Sankey diagram down to the stacked bar graph, we reduced the padding to 0 and also changed the layout to 0. The 0 layout will draw the objects from the top of the SVG canvas and build downwards. In order to draw an SVG shape, first a canvas must be specified or created if not existing. Shapes are then drawn on top of the canvas just like the process of any brush and canvas physical painting. In order to base the overall graphic along the bottom of the SVG canvas, we utilize the SVG transform attribute, which controls positioning, to flip the SVG canvas through the use of two specific inputs: translate (x and y movement) and scale (inverting/flipping image). With this positioning adjustment, the position of each of the major elements entering the screen also needs to be transformed.

These positioning changes to the default library added a few advantages to the public examples. By reducing padding and basing all bars to the bottom of the canvas, we can better compare respective cluster sizes over time as opposed to the default layouts which may require repositioning of nodes though click and drag interactions. Along with the use of a consistent color scheme, the proposed visual quickly shows relatively consistent proportions to the four main player groups at each interval. Also, the entire bars represent the total player count over time, with which we see a diminishing trend outside of a slight resurgence in the last few months. Another key point of interest, the difference between incoming and outgoing players can be seen by the lack of links connecting to nodes at particular points. More specifically, if there is space on the left side of a node that does not have a link connecting to it; this means that there were more incoming players than outgoing. Across all clusters, at any given month there is inflow and outflow to each unique cluster, albeit having a small proportion in many cases.

Finally, we utilize a second screen to help drill down into the history of a cluster. On a mouse click of any given node, we remove the links, and transition to a stacked bar graph with d3.layout.stack(). The resulting chart shows which cohorts the players from the selected group belonged to in preceding and proceeding months. The y-axis will scale appropriately based on the relative size of the player group being selected.

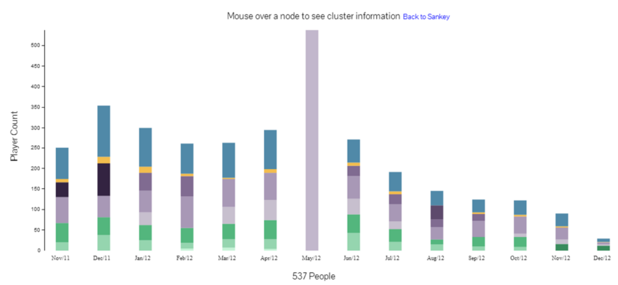

Figure 4: Visualization on cluster click through: clicking on a node reveals the distribution that lead up to the node and the distribution in proceeding months for said node.

7. Key results

Due to space constraints, only a few key results will be outlined here. Firstly, the Sankey diagram (http://powerful-meadow-8588.herokuapp.com/) allows the observation of variations in player behavior over time. This can serve as benchmark for tracking disaffected players, compromised accounts, which types of players are at the highest risk for leaving the game, etc. While the process for our research was manual, our methodology can be scaled up to be fully automated. Secondly, players do not remain in single cluster for the duration of their lifetime. Players move relatively freely across clusters from month to month.

Thirdly, taking an in-depth look at the individual cluster nodes, players are most likely to enter and leave the game as a casual player, resulting in the highest inflow/outflow as a percentage of the cluster compared to other clusters. We can speculate that this is where players first get acclimated to the auction Hardcore players, on the other hand showed the highest proportion of consistency across clusters. At any given month, around half of hardcore players remained Hardcore players in the preceding and proceeding months.

The insularity of the segment was higher than that of any other group. Additionally, hardcore players were least likely to leave the game. Especially in the early months of data tracking, Hardcore auction users and forum posters were most likely to stick with the game until the end. Hardcore players stayed in the game μ = 5.2 months, Forum posters μ = 7.4 months, while Casual users μ = 3.2 months. This highlights the importance of users' engagement. Those who were active in the online community outside of the game had the highest retention rates. Moving from Hardcore to Casual status can thus be an indication of potential player departure.

8. Discussion and conclusion

The analysis presented here focuses on the auction house component of the economic system of *Glitch*, and combines with temporal perspectives and the use of Sankey diagrams to track and visualize the flow of players between behavioral clusters. It is clear from the raw numbers that *Glitch* experienced a declining user base over time, but by integrating analysis with visualization, it is possible to observe the user base and auction house behaviors change over time, providing a dashboard for tracking the temporal elements of the clusters and associated behaviors.

The results show that the auction house of *Glitch* was used by the majority of the players, and this use followed a limited set of patterns, as exemplified by the clusters. The fundamental methodological approach presented here is based on individual components that are well known - k-means clustering of behavioral data from games (Drachen et al. 2012), temporal bins (Sifa et al. 2013) and Sankey diagrams (Riehman et al. 2005; Schmidt 2008). However, by combining these, a framework is established that is useful across a wide variety of progression-based or temporal-based scenarios, providing a useful tool for monitoring the overall health of games. Strong shifts in the data, either from players moving across clusters over time or player departure can indicate an issue that needs to be resolved or a disaffected group that is no longer satisfied with the game. Alternate uses of the framework include evaluating playstyles across multiple game levels, with the goal of examining which strategies lead players to complete a game. Other methods for segmenting players, notably funnel analysis (Fields and Bohannon 2011), can be adapted for temporal bins but do not involve any analysis of gameplay - only player numbers surviving each progressive step in the funnel.

Future work will extend the current framework to other problem areas in game analytics and integrate multiple clustering algorithms Finally, on the visualization front, a next step is rewriting the Sankey.js plugin code to enable acceptance of per player data as raw input would allow the construction of a system that enable drill-down analysis down to the user level, not just the level of individual clusters and time bins.

References

Bartle, R. A. 2003. *Designing Virtual Worlds*. New Riders.

Bloomfield, R. J.; Cho, Y-J. 2011. Unregulated Stock Markets in Second Life. *Southern Economic Journal.* 78.1

Bostock, M. 2012 Sankey Diagrams. URL: http://bost.ocks.org/mike/Sankey/

Castronova, E., Williams, D., Shen, C., Ratan, R., Xiong, L., Huang, Y. and Kegan, B. 2009. As real as real? Macroeconomic behavior in a large-scale virtual world. *New Media Society* 11:685

Drachen, A. Canossa, A. and Yannakakis, G. N. 2009. Player Modeling using Self-Organization in Tomb Raider: Underworld. In *Proc. of the IEEE Symposium on Computational Intelligence and Games,* 1–8 (Milan, Italy).

Drachen, A., Sifa, R., Bauckhage, C. and Thurau, C. 2012. Guns, Swords and Data. In *Proc. of the IEEE Symposium on Computational Intelligence and Games,* 163-170 (Granada, Spain).

Fields, T. and Cotton, B. 2011. *Social Game Design: Monetization Methods and Mechanics.* Morgan Kaufmann.

Grabianowski, I. 2013. *Real Economics Studies Virtual Economy in EVE Online.* iO9. URL: http://io9.com/5057849/real-economist-studies-virtual-economy-in-eve-online

Keegan, B.; Ahmed, M. A.; Williams, D.; Srivastava, J. and Contractor, N. 2010. Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games. In *Proc. IEEE 2nd International Conference on Social Computing.*

Landwehr, P. M. 2013. *A Collection of Economic and Social Data from Glitch, a Massively Multiplayer Online Game*. Technical Report CMU-ISR-13-100. Carnegie Mellon University. URL: http://reports-archive.adm.cs.cmu.edu/anon/isr2013/abstracts/13-100.html

Lohdonvirta, V. 2009. Virtual item sales as a revenue model: identifying attributes that drive purchase decisions. Electronic Consumer Research, 9: 97-113.

Medler, B. 2012. Play with Data – An Exploration of Play Analytics and its Effect on Player Experiences. PhD-Thesis. Georgia Institute of Technology, School of Literature, Communication and Culture.

Morrison, M. and Fontenla, M. 2013. Price Convergence in an online virtual world. Empirical Economics, 44:1053-1064.

Plumer, B. The economics of video games. *The Washington Post*. September 28th, 2012.

Riehmann, R.; Hanfler, M. and Froehlich, B. 2005. Interactive Sankey Diagrams. In *Proc. IEEE INFOVIS symposium*, 233-240.

Schmidt, M. 2008. The Sankey Diagram in Energy and Material Flow Management. *Journal of Industrial Ecology* 12:1, 82-94.

Seif El-Nasr, M., Drachen, A. and Canossa, A. 2013. *Game Analytics – Maximizing the Value of Player Data*. Springer.

Shokrizade, R. 2013. How I Used EVE Online to Predict the Great Recession. *Gamasutra*. URL: http://gamasutra.com/blogs/RaminShokrizade/20130405/189984/How_I_Used_EVE_Online_to_Predict_the_Great_Recession.php

Sifa, R.; Drachen, A. and Bauckhage, C. 2013. Evolutionary Behavioral Clustering in Tomb Raider. In Proc. IEEE Computational Intelligence in Games (Granada, Spain).

Simpson, Z. B. 1999. The In-game Economics of Ultima Online. Presented at *Game Developers Conference* (San Jose, CA).

Tufte, E. 1983. *The Visual Display of Quantitative Information.* Graphics Press.

Varoufakis, Y. 2012. It all Began with a Strange Email. *Valve Economics.* URL: http://blogs.valvesoftware.com/economics/it-all-began-with-a-strange-email/