On Churn and Social Contagion

Zoheb Borbora
Dept. of Computer Science
University of Minnesota
Minneapolis, USA
borbo001@umn.edu

Arpita Chandra
Center for Exact Humanities
International Institute of Information Technology
Hyderabad, India
arpita.chandra@research.iiit.ac.in

Ponnurangam Kumaraguru *
Dept. of Computer Science
Indraprastha Institute of Information Technology
Delhi, India
pk@iiitd.ac.in

Jaideep Srivastava
Dept. of Computer Science
University of Minnesota
Minneapolis, USA
srivasta@cs.umn.edu

***Abstract*—Massively Multiplayer Online Role-Playing Games (MMORPGs) are persistent virtual environments where millions of players interact in an online manner. We study the problem of player churn and social contagion using MMORPG game logs by analyzing the impact of a node's churn behavior on its immediate neighborhood or group. The two key research questions in this paper are - When an active node, *ego*, becomes dormant, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, 1) based on *ego's* characteristics and *ego's* relationship with *alter* and 2) based on the activity behavior of *alter's* remaining neighbors. We use a supervised learning framework to study the impact of player churn and social contagion. Experimental results show that the classification models perform substantially better than random for both the research problems. Finally, we use a data-driven approach to propose a player typology based on degree of socialization and analyze churn behavior among these player types. Experimental results show that the *loner* player type is much more likely to churn than the *socializer* player types and as the degree of socialization decreases among socializers, the propensity to churn increases.**

***Index Terms*—churn, social contagion, supervised learning, clustering, player typology**

I. Introduction

Most of existing churn research have focused on modeling individual churn behavior without considering network effects. Recent work have started to investigate the influence of social ties in individual churn analysis [2], [3]. However, all the studies so far have focused on the impact on individual churn based on neighbors who have already churned. In this paper, we study what is the impact on an actor's neighborhood if the actor churns from the network. To the best of our knowledge, this question has not been studied in the churn literature and we believe this would be our novel contribution to the area.

An important aspect of any churn prediction model is how early the model can identify that an actor is at risk of churn. Typical churn prediction models would predict whether an actor is likely to churn in the next time window; which could be a week or month. However, at this point it may already be too late for Customer Relationship Managers to take effective counter-measures to prevent the churn from happening. Prior studies have shown that socialization has an impact on individual churn [1]–[3]. We can therefore make the assumption that when an actor churns, it has an impact on the actor's neighborhood. Specifically, we would expect the churn likelihood of the actor's neighbor to change and the magnitude of the change would depend on the degree and nature of influence that the actor has on the neighbor. This can provide early warning regarding an actor's propensity to churn rather than just doing next-interval prediction. *Key contributions* of the paper are listed below.

First, we address the question - When an active node, *ego*, becomes dormant, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on *ego's* characteristics and *ego's* relationship with *alter*? We consider individual game-based and node-based features for *ego* and also features based on the existing relationship between *alter* and *ego*. Results indicate that *ego's* centrality/prestige in the network is a key determinant of *alter's* activity behavior after *ego* becomes dormant. *Ego's* character level, which is indicative of expertise level, is also a key factor in *alter's* change in behavior. Among the features based on the existing relationship between *alter* and *ego*, we find that the number of common neighbors and the adar-adamic index are key determinants in the contagion process. Finally, results indicate that homophily-based features between *alter* and *ego* are not very discriminating in predicting dyadic influence.

Second, we address the question - When an active node, *ego*, becomes dormant, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on the activity

*Work was done during his sabbatical at IIIT Hyderabad

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08?/$15.00
http://dx.doi.org/10.1145/3341161.3344378

behavior of *alter's* remaining neighbors? Results indicate that *alter's* behavior is strongly impacted by the number of remaining active neighbors and the strength of the relationship with those neighbors. Thus, we find that there is a strong social influence in effect wherein *alter's* activity behavior is impacted by the activity behavior of the players around *alter*. This is in keeping with existing models of diffusion in the literature [4].

Third, we use a supervised learning framework to study the impact of player churn and social contagion. Experimental results show that the classification models perform substantially better than random for both the research problems.

Fourth, we use a data-driven approach to propose a player typology based on degree of socialization of players and analyze churn behavior among these player types. Experimental results show that the *loner* player type is much more likely to churn than the *socializer* player types and as the degree of socialization decreases among socializers, the propensity to churn increases.

To the best of our knowledge, the problem of player churn and social contagion has not been studied in the literature and in general, the problem of social contagion has not been studied in a supervised learning framework.

II. Background and Related Work

In this paper, we focus on the impact of a node's churn behavior on its immediate neighborhood or group - the underlying sociological processes of this study relate to the fields of social influence, social contagion and group dynamics. Specifically, the research questions in this paper relate to behavioral contagion [5] with the behavior in question being a player's churn in an online game setting.

Recent work has started looking at social contagion and behavior cascades in human social networks [6], [7] - this encompasses a diverse spectrum of behavior and affective states such as happiness [8], loneliness [9], depression [10], smoking [11], alcohol consumption [12] and even obesity [13]. Aral and Walker used in vivo randomized experimentation to identify influence and susceptibility in networks using a sample of 1.3 million Facebook users [14]. In a separate study, Aral et al. [15] distinguished between contagion and homophily [16] effects in a dataset documenting product adoption in a large network.

Traditionally, models of diffusion by which information, ideas and influence spread through a network has been studied extensively in a number of domains such as diffusion and adoption of innovations [17], [18], spread of infectious diseases in epidemiology [19], [20], the effects of "word of mouth" and "viral marketing" in the promotion of new products [21]–[23], trust propagation through networks [24], [25] and more recently, information diffusion in online social networks [26]–[29].

We study behavioral contagion in small groups within online games in this paper. To the best of our knowledge, no other work has studied this problem. Previous studies have only looked at in-game player organizations in MMORPGs called *guilds*. Williams et al. did an extensive study of the social life

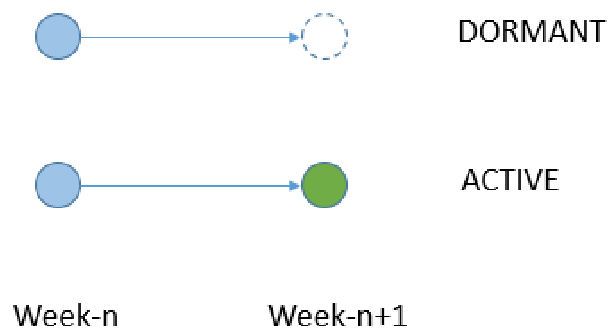

Fig. 1. Dormancy - Classification Labels

of guilds in *World of Warcraft* by interviewing a representative sample of players in the game [30]. Ducheneaut et al. analyzed the structural properties of guilds in *World of Warcraft*, and examined some of the factors that could explain the success or failure of these groups [31]. Kang et al. [32] studied the social dynamics of a guild in its early stages and found that social activities that lead to an increase in an individual's profit are the main factors in strengthening and expanding the lifetime of the guilds in these stages, more so than pure friendship and loyalty, especially in the very early stages. Patil et al. [33] used a classification-based approach to predict guild stability and constructed a range of features that describe group composition, group structure and group activities.

III. Methodology

We use a supervised learning approach to study the impact of *ego's* churn on its immediate neighborhood or group.

A. Classification Labels

A node is labeled based on whether or not it becomes dormant going from one week to the next. Figure 1 illustrates the labelling scheme.

- **DORMANT**: A node is labeled as DORMANT in week n if its session length drops to zero in week $n + 1$.
- **ACTIVE**: A node is labeled as ACTIVE in week n if its session length has a non-zero value in week $n + 1$.

B. Dyadic Influence of ego on alter's activity behavior

The first research question we address is

RQ1: When an active node, *ego*, becomes DORMANT, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on *ego's* characteristics and *ego's* relationship with *alter*?

The purpose of this study is to gain insight into the factors that influence *alter's* behavior when *ego* is no longer in the network, as illustrated in Figure 2 where *ego* becomes DORMANT in week $n + 1$. In this particular scenario, *ego* has three neighbors in week n. We consider individual game-based and node-based features for *ego* in week n and also features of the edges *ego*-A, *ego*-B and *ego*-C in week n in order to predict the label for A, B and C in week n+1.

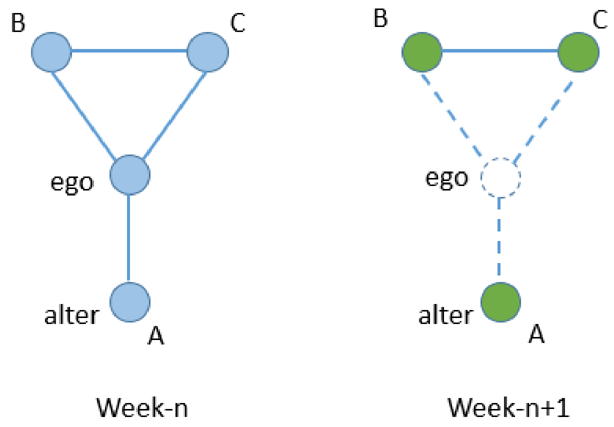

Fig. 2. Dyadic Influence of *ego* on *alter* - illustrative example

1) Game-based features for ego*:* These features are intended to capture *ego's* engagement and experience in the game. A key indicator of engagement is the amount of playtime *ego* puts into the game and this is captured using session-related features. The activity logs available to us only record player actions and did not explicitly define sessions. So, we used a simple heuristic to define player-sessions - a session consists of sets of activities which are separated by no more than 30 minutes. We consider *ego's* overall sessions and sessions spent in groups. While the *Overall Session* features give a measure of *ego's* overall engagement, the *Group Session* features are a measure of *ego's* degree of socialization/group-oriented activities during the week. The features in this category are listed below -

- *Overall Number of Sessions* for *ego* during the week.
- *Overall Session Length (mins)* for *ego* during the week.
- *Overall Inter-Session Length (mins)*: Time between sessions for *ego* during the week.
- *Number of Sessions in Groups* spent by *ego* during the week.
- *Session Length (mins) in Groups* spent by *ego* during the week.
- *Inter-Session Length (mins) in Groups* spent by *ego* during the week.
- *Player Character Level* - max character level for *ego* during the week. This feature is intended to capture *ego's* level of expertise in the game. A player with greater expertise in the game can be expected to have greater influence on its immediate neighbors.
- *Experience Points* collected by *ego* during game play in the week. This feature is intended to capture *ego's* degree of success in game-play activities during the week. Successful players can be expected to have greater influence on its immediate neighbors in terms of motivating them towards game-play.

2) Network-based features for ego*:* These features are intended to capture the importance/prestige of *ego* in the network. The features were were computed using the Pajek [34] program for analysis and visualization of large networks.

- *Degree Centrality*: Degree centrality is based on the number of edges incident upon the node (i.e., the number of ties that a node has) [35]

$$C_D(ego) = Degree(ego) \tag{1}$$

- *Closeness Centrality*: The *farness* of a node s is defined as the sum of its distances to all other nodes, and its *closeness* is defined as the reciprocal of the farness. Thus, the more central a node is the lower its total distance to all other nodes.
- *Betweenness Centrality*: Betweenness centrality of a node is a measure based on the number of shortest paths that a vertex lies in [36]

$$C_B(ego) = \sum_{s \neq i \neq t \in V} \frac{\sigma_{st}(ego)}{\sigma_{st}} \tag{2}$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(i)$ is the number of shortest paths from s to t that pass through vertex ego.
- *Clustering Coefficient*: Clustering Coefficient is the fraction of pairs of a person's collaborators who have also collaborated with one another [37].

$$C(ego) = \frac{3 \times \textit{number of triangles}}{\textit{number of connected triples}} \tag{3}$$

3) Edge-based features for ego *and* alter*:* These features are intended to capture various facets of the existing relationship between *ego* and *alter* such as a) shared sessions, b) common neighbors in the network and c) homophily. Shared sessions between *alter* and *ego* are a good indicator of the strength of interpersonal connection between the nodes. Features based on common neighbors and similarity metrics are intended to capture . Homophily-based features are intended to capture similar tastes, preferences and demographics between *alter* and *ego*.The features in this category are listed below -

- *Shared Number of Sessions* between *alter* and *ego* during the week.
- *Shared Session Length (mins)* between *alter* and *ego* during the week.
- *Common Neighbors*: as given by

$$n(ego, alter) = |\mathbb{G}(ego) \cap \mathbb{G}(alter)| \tag{4}$$

- *Jaccard Index* [38]: a statistic used for comparing the similarity and diversity of sets.

$$\gamma(ego, alter) = \frac{|\mathbb{G}(ego) \cap \mathbb{G}(alter)|}{|\mathbb{G}(ego) \cup \mathbb{G}(alter)|} \tag{5}$$

- *Adar-Adamic Index*: Neighbors with few connections have more weight in capturing the similarity of nodes i and j [39].

$$\alpha(ego, alter) = \sum_{k \in \mathbb{G}(ego) \cap \mathbb{G}(alter)} \frac{1}{\log(|\mathbb{G}(k)|)} \tag{6}$$

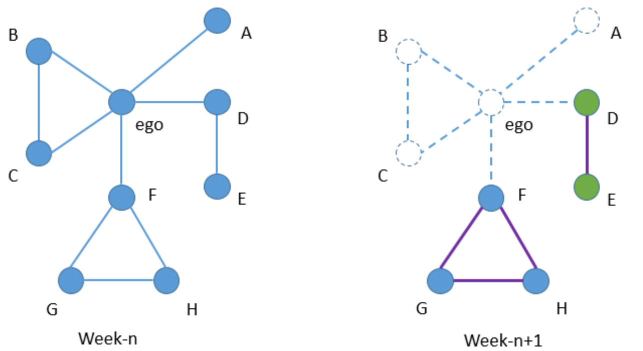

Fig. 3. Neighborhood Influence on *alter* - illustrative example

- *Character Race Homophily Indicator*: identifies whether *alter* and *ego* belong to the same race within the game.
- *Character Class Homophily Indicator*: identifies whether *alter* and *ego* belong to the same class within the game. In a generalized setting, this could be interpreted as whether the two actors have similar skills and interests.
- *Character Guild Homophily Indicator*: identifies *alter* and *ego* to the same guild in the game. In a generalized setting, this could be interpreted as whether the two actors belong to the same organization or group and hence share a common goal.
- *Character Gender Homophily Indicator*: identifies *alter* and *ego* have the same character gender in the game.
- *Real Country Homophily Indicator*: identifies whether *alter* and *ego* belong to the same real-world country.

C. *Neighborhood Influence on* alter's *activity behavior*

The second research question we address is

RQ2: When an active node,*ego*, becomes DORMANT, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on the activity behavior of *alter's* remaining neighbors?

The purpose of this study is to gain insight into the factors that cause *alter* to leave or stay after *ego* has left the network, as illustrated in Figure 3 where *ego* becomes DORMANT in week $n+1$. In this particular scenario, *ego* has five neighbors in week n and depending on the remaining neighbors, each of these *alter* may undergo a different change in their activity behavior. For example, B and C may become DORMANT as well since they were part of a triad with *ego*. D might continue to remain ACTIVE because of E's influence. F, G and H maybe relatively unaffected because they are part of a stable triad.

In order to study this question, we look at the neighbors for *alter* for the week n. The general idea is that because of social influence [4], a node's behavior will primarily be impacted by the behavior of the players around *alter*. The specific features are listed below -

- *Number of DORMANT neighbors*: for *alter* in the current week.
- *Number of ACTIVE neighbors*: for *alter* in the current week.

TABLE I
WEEKLY ACTIVITY NETWORKS - BREAKDOWN BY DORMANCY LABELS

Ego Label	**Week** 14	15	16	17
DORMANT	1266 (21.69%)	1308 (22.16%)	1488 (24.23%)	1480 (24.69%)
ACTIVE	4572 (78.31%)	4594 (77.84%)	4652 (75.77%)	4515 (75.31%)
Total Nodes	5838	5902	6140	5995
Total Edges	35713	35592	38114	37928

TABLE II
EQII WEEKLY GROUP NETWORKS - AVERAGE DEGREE BY DORMANCY LABELS

Ego Label	**Week** 14	15	16	17
DORMANT	4.92	5.21	5.08	5.03
ACTIVE	14.22	13.99	14.75	15.09

- *Edge-weighted DORMANT neighbors*: for *alter* in the current week.
- *Edge-weighted ACTIVE neighbors*: for *alter* in the current week.

D. *Player typology based on group interactions*

Finally, we use a data-driven approach to segment players into different types based on their degree of socialization and analyze churn behavior among these player segments. We use the following features to segment the nodes into different player types.

- *Number of neighbors*: in the current week.
- *Fraction of group sessions*: in the current week, given by

$$\frac{\textit{Number of group sessions}}{\textit{Total sessions}} \tag{7}$$

- *Average tie strength*: in the current week, given by

$$\frac{\textit{Number of group sessions}}{\textit{Number of neighbors}} \tag{8}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. *Data Description*

Data from the MMORPG - Sony Everquest II [1] was used for the experiments and analysis. The game data had four servers and in-game relationships among players include housing (player grants access to his/her house to another player), mentoring (experienced player mentoring another player to help the latter acquire skills and level up within the game), trading (buying and selling of in-game items), chat and grouping. As a player levels up within the game, the quests become progressively more difficult and it becomes essential for players to play in groups to accomplish tasks and progress

[1]https://www.everquest2.com/

TABLE III
DYADIC INFLUENCE DATASET

Alter Label (Week 14)	**Sample Size**	**Percentage**
DORMANT	1228	19.7%
ACTIVE	5005	80.3%
Total	6233	100%

through the game [2]. For the purposes of our analysis, we used data from the *guk* server and the grouping relationships to build networks across the weeks.

Table I gives a breakdown of the nodes and edges in the group networks across 4 weeks for the two labels - DORMANT and ACTIVE. For example, there are a total of 5838 nodes and 35713 undirected edges in the week 14 group network. Of the 5838 nodes in week 14, 1266 become DORMANT (session length goes to zero) and 4572 remain ACTIVE (non-zero session length). We observe that, on average, around 23.19% of the nodes become DORMANT and 76.81% of the nodes remain ACTIVE going into the next week.

Table II shows the average degree of nodes for the two labels. We observe that, on average, DORMANT nodes have 5.06 neighbors and ACTIVE nodes have 14.52 neighbors. This indicates that nodes which become dormant have a disproportionately smaller number of neighbors.

B. *Experiment 1 - Dyadic Influence on* alter's *activity behavior*

In this experiment, we use a binary classification approach to study the factors that impact activity of *ego's* immediate neighbor, *alter*, when *ego* becomes DORMANT - based on *ego's* characteristics and *ego's* relationship with *alter*. The week 14 group network is used for these purposes. There are 1266 nodes in the week 14 network which become DORMANT and there are 6233 edges connected to these nodes. Thus, we get 6233 samples for the experiment (refer Table III). Of these samples, *alter* itself becomes DORMANT in 1288 (19.7%) of the samples and *alter* remains ACTIVE in 5005 (80.3%) of the samples.

Table IV lists the 22 features (refer Section III-B) in decreasing order of their info-gain value. Based on this table, we observe that the top five features in predicting change in *alter's* behavior after *ego* has become DORMANT are -

- *Ego's closeness centrality*: Closeness centrality of a node is a centrality measure based on how close the node is to all other nodes. The importance of this feature indicates that ego's location within the network is indicative of its social influence on its neighbors and is highly discriminating in terms of determining *alter's* behavior when *ego* leaves.
- *Number of common neighbors between alter and ego*: Common neighbors between *ego* and *alter* are an indicator of the strength of the connection between *ego* and *alter*.
- *Ego's character level*: This feature indicates that *ego's* level of expertise is important in determining *alter's* behavior.
- *Adar-adamic index for alter and ego*: Adar-adamic index between two nodes is greater when common neighbors of the two nodes have fewer connections. The importance of this feature indicates that when the common neighbors between alter and ego are primarily connected to them, the influence is stronger.
- *Ego's degree centrality*: This feature indicates that *ego's* connectivity, in general, is indicative of its influence on *alter*.

Furthermore, we observe that homophily-based features between *alter* and *ego* are not very discriminating in predicting dyadic influence.

Table V shows the precision, recall and f-measure results for the dyadic influence model on the DORMANT and ACTIVE labels. We observe that all the models do much better than random for both the labels (refer Table III).

C. *Experiment 2 - Neighborhood Influence on* alter's *activity behavior*

In this experiment, we use a binary classification approach to study the factors that impact activity of *ego's* immediate neighbor, *alter*, when *ego* becomes DORMANT -based on the activity behavior of *alter's* remaining neighbors. The week 14 group network is used for these purposes. There are 1266 nodes in the week 14 network which become DORMANT and there are 3157 unique *alter's* connected to these nodes. Thus, we get 3157 samples for the experiment (refer Table VI). Of these samples, *alter* itself becomes DORMANT in 770 (24.39%) of the samples and *alter* remains ACTIVE in 2387 (75.61%) of the samples.

Table VII lists the 4 neighbor based features (refer Section III-C) in decreasing order of their info-gain value. Based on this table, we observe that *alter's* behavior is strongly impacted by the number of remaining active neighbors and the strength of the relationship with those neighbors. This is indicative of a social influence effect going on.

Table VIII shows the precision, recall and f-measure results for the neighborhood influence model on both the labels. We observe that the models do much better than random for the both the labels (refer Table VI).

D. *Experiment 3 - Player types and churn behavior*

K-means clustering is used to segment the Week 14 dataset into different clusters. In order to get an idea for the number of naturally occurring clusters in the dataset, we plot the Within Cluster Sum of Squared Errors (SSE) against the number of clusters, k - (refer Figure 4). To get an estimate of the optimal number of clusters, one might look for the knee of the curve. In this case, we observe that optimal number of such clusters is *4*. Quantitative techniques may also be used for this sort of analysis. For example, the Calinski-Harabasz index (CH) [40]

[2]http://eq2.wikia.com/wiki/Groups

TABLE IV
DYADIC INFLUENCE MODEL - INFO-GAIN RANKING OF FEATURES

Info-gain Ranking	Feature	Info-gain Value
1	ego_curr_week_closeness_cent	0.0856
2	alter_ego_curr_week_common_neighbors	0.07826
3	ego_curr_week_pc_level	0.06883
4	alter_ego_curr_week_adar_adamic_index	0.0638
5	ego_curr_week_degree_cent	0.05962
6	alter_ego_curr_week_shared_num_sessions	0.04531
7	ego_curr_week_clustering_coef	0.03559
8	ego_curr_week_total_xp	0.02996
9	alter_ego_guild_indicator	0.02622
10	alter_ego_country_indicator	0.02337
11	alter_ego_race_indicator	0.02331
12	alter_ego_gender_indicator	0.02202
13	alter_ego_class_indicator	0.02127
14	ego_curr_week_sl_mins	0.0143
15	ego_curr_week_betweenness_cent	0.01413
16	ego_curr_week_group_sl_mins	0.01385
17	ego_curr_week_num_group_sessions	0.0137
18	ego_curr_week_num_sessions	0.01357
19	alter_ego_curr_week_shared_sl_mins	0.00948
20	ego_curr_week_isl_mins	0.00557
21	ego_curr_week_group_isl_mins	0.00257
22	alter_ego_curr_week_jaccard_index	0

TABLE V
DYADIC INFLUENCE MODEL - 10-FOLD CROSS VALIDATION RESULTS

	DORMANT			ACTIVE		
Model	**Precision**	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
J48	54.95	36.64	43.97	85.63	92.63	88.99
Logistic	62.63	23.62	34.3	83.74	96.54	89.69
kNN	46.6	32.41	38.23	84.57	90.89	87.62
Naive Bayes	36.6	55.7	44.17	87.53	76.32	81.55
Perceptron	60.54	23.86	34.23	83.74	96.18	89.53
AdaBoost	53.02	40.8	46.11	86.25	91.13	88.62
Bagging	63	38.27	47.62	86.19	94.49	90.14

TABLE VI
NEIGHBORHOOD INFLUENCE DATASET

Alter Label (Week 14)	**Sample Size**	**Percentage**
DORMANT	770	24.39%
ACTIVE	2387	75.61%
Total	3157	100%

evaluates the cluster validity based on the average between- and withincluster sum of squares

Table IX lists the cluster centroids from the output of the *k-means* algorithm - we identify the following four player types based on characteristics of the cluster that the player belongs

- *Social Butterfly*: This player type interact with a large number of other players (mean of 115.04 neighbors) but the average tie strength is quite low (mean of 0.29) with the other players. In other words, such players have a lot of weak connections
- *Large Pack Wolf*: This player type interact with a smaller number of other players as compared to a *Social Butterfly* (mean of 58.86 neighbors) but the average tie strength is more that of a *Social Butterfly* (mean of 0.41).
- *Small Pack Wolf*: This player type interact with a smaller number of other players as compared to a *Large Pack Wolf* player (mean of 26.35 neighbors) but the average tie strength is more that of a *Large Pack Wolf* player (mean of 0.87).
- *Lone Wolf*: This player type interacts with the small number of other players as compared to the other player types

TABLE VII
NEIGHBORHOOD INFLUENCE MODEL - INFO-GAIN RANKING OF FEATURES

Info-gain Ranking	Feature	Info-gain Value
1	num_curr_week_neighbors_ACTIVE	0.1881
2	weighted_curr_week_neighbors_DORMANT	0.1437
3	weighted_curr_week_neighbors_ACTIVE	0.1265
4	num_curr_week_neighbors_DORMANT	0.0167

TABLE VIII
NEIGHBORHOOD INFLUENCE MODEL - 10-FOLD CROSS VALIDATION RESULTS

	DORMANT			ACTIVE		
Model	**Precision**	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
J48	65.83	44.29	52.95	83.74	92.58	87.94
Logistic	65.32	42.08	51.18	83.24	92.79	87.76
kNN	54.22	48.44	51.17	83.93	86.8	85.34
Naive Bayes	54.63	60.52	57.42	86.81	83.79	85.27
Perceptron	63.93	43.51	51.78	83.48	92.08	87.57
AdaBoost	60.74	47.01	53	84.07	90.2	87.03
Bagging	66.54	45.19	53.83	83.98	92.67	88.11

(mean of 6.82 neighbors) but the average tie strength is highest compared to the rest (mean of 2.63). In other words, such players have a very few strong connections

Among the three player types, we observe that the fraction of group sessions is in the majority (0.6 to 0.67) for the first three categories but in the minority (0.46) for the *Lone Wolf* category.

Finally, we look at the proportion of churners and non-churners for each player type (refer Table X). We observe that as we go down the list of player types in decreasing order of neighbors and fraction of group sessions and increasing order of average tie strength, the propensity to churn increases i.e *Social Butterfly* > *Pack Wolf* > *Lone Wolf*. This would indicate that the loner player type is much more likely to churn than the socializer player types and as the degree of socialization decreases among socializers, the propensity to churn increases.

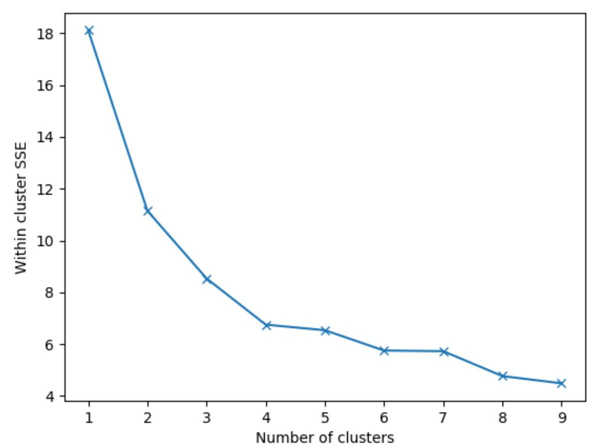

Fig. 4. K-means for getting player types

V. CONCLUSION AND FUTURE WORK

We address two research questions related to player churn and social contagion in this paper. We used a supervised learning framework for the study and found that the classification models perform substantially better than random for both the research problems.

First, we address the question - When an active node, *ego*, becomes dormant, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on *ego's* characteristics and *ego's* relationship with *alter*? We consider individual game-based and node-based features for *ego* and also features based on the existing relationship between *alter* and *ego*. Results indicate that *ego's* centrality/prestige in the network is a key determinant of *alter's* activity behavior after *ego* becomes dormant. *Ego's* character level, which is indicative of expertise level, is also a key factor in *alter's* change in behavior. Among the features based on the existing relationship between *alter* and *ego*, we find that the number of common neighbors and the adar-adamic index are key determinants in the contagion process. Finally, results indicate that homophily-based features between *alter* and *ego* are not very discriminating in predicting dyadic influence.

Second, we address the question - When an active node, *ego*, becomes dormant, what is the impact on the activity behavior of *ego's* immediate neighbor, *alter*, based on the activity behavior of *alter's* remaining neighbors? Results indicate that *alter's* behavior is strongly impacted by the number of remaining active neighbors and the strength of the relationship with those neighbors. Thus, we find that there is a strong social influence in effect wherein *alter's* activity behavior is impacted by the activity behavior of the players around *alter*. This is in

TABLE IX
PLAYER TYPES BASED ON K-MEANS CLUSTERING

Player Label	No. of samples	Cluster centroids		
		Number of neighbors	**Fraction of group sessions**	**Average Tie Strength**
Social Butterfly	137	115.04	0.67	0.29
Large Pack Wolf	549	58.86	0.64	0.41
Small Pack Wolf	1223	26.35	0.6	0.87
Lone Wolf	2706	6.82	0.46	2.63

TABLE X
CHURN BEHAVIOR BASED ON PLAYER TYPES

Player Type	Churners	Non-churners
Social Butterfly	7 (5.1%)	130 (94.9%)
Large Pack Wolf	48 (8.7%)	501 (91.3%)
Small Pack Wolf	220 (18%)	1003 (82%)
Lone Wolf	1176 (43.5%)	1530 (56.5%)

keeping with existing models of diffusion in the literature [4].

Finally, we use a data-driven approach to propose a player typology based on degree of socialization of players and analyze churn behavior among these player types. Experimental results show that the *loner* player type is much more likely to churn than the *socializer* player types and as the degree of socialization decreases among socializers, the propensity to churn increases.

The research questions and analysis in this paper deal with gaining insights into the churn contagion process and discovering the network substructures which facilitate or inhibit the contagion process. We believe understanding these factors can be quite helpful in early identification of at-risk players and give the CRM folks a window of opportunity to address and mitigate the risk. One possible action that CRM folks could take would be to recommend friends and groups to players who are at risk of churn.

REFERENCES

[1] Borbora, Zoheb Computational Analysis of Churn in Multiplayer Online Games. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/181699, 2015

[2] Koustuv Dasgupta, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukherjea, Amit A. Nanavati, and Anupam Joshi. Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '08, pages 668–677, New York, NY, USA, 2008. ACM.

[3] Jaya Kawale, Aditya Pal, and Jaideep Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, pages 423–428, Washington, DC, USA, 2009. IEEE Computer Society.

[4] Mark Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[5] Ladd Wheeler. Toward a theory of behavioral contagion. *Psychological review*, 73(2):179–192, 1966. ID: 1966-05303-001.

[6] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[7] James H Fowler and Nicholas A Christakis. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12):5334–5338, 2010.

[8] James H Fowler, Nicholas A Christakis, et al. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Bmj*, 337:a2338, 2008.

[9] John T Cacioppo, James H Fowler, and Nicholas A Christakis. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology*, 97(6):977, 2009.

[10] J Niels Rosenquist, James H Fowler, and Nicholas A Christakis. Social network determinants of depression. *Molecular psychiatry*, 16(3):273–281, 2011.

[11] Nicholas A Christakis and James H Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

[12] J Niels Rosenquist, Joanne Murabito, James H Fowler, and Nicholas A Christakis. The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine*, 152(7):426–433, 2010.

[13] Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[14] Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.

[15] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

[16] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001, http://www.annualreviews.org/doi/pdf/10.1146/annurev.soc.27.1.415.

[17] J.S. Coleman, E. Katz, H. Menzel, and Columbia University. Bureau of Applied Social Research. *Medical innovation: a diffusion study*. Advanced Study in Sociology. Bobbs-Merrill Co., 1966.

[18] E.M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 2003.

[19] N.T.J. Bailey. *The Mathematical Theory of Infectious Diseases and Its Applications*. Mathematics in Medicine Series. Griffin, 1975.

[20] R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford science publications. OUP Oxford, 1992.

[21] Jacqueline J. Brown and Peter H. Reingen. Social Ties and Word-of-Mouth Referral Behavior. *The Journal of Consumer Research*, 14(3):350–362, 1987.

[22] Vijay Mahajan, Eitan Muller, and Frank M. Bass. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1):pp. 1–26, 1990.

[23] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 57–66, New York, NY, USA, 2001. ACM.

[24] C.-N. Ziegler and G. Lausen. Spreading activation models for trust propagation. In *e-Technology, e-Commerce and e-Service, 2004. EEE '04. 2004 IEEE International Conference on*, pages 83 – 97, march 2004.

[25] Elizabeth Gray, Jean-Marc Seigneur, Yong Chen, and Christian Jensen. Trust propagation in small worlds. In Paddy Nixon and Sotirios Terzis, editors, *Trust Management*, volume 2692 of *Lecture Notes in Computer Science*, pages 1072–1072. Springer Berlin / Heidelberg, 2003.

[26] Jon Kleinberg. Bursty and hierarchical structure in streams. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 91–101, New York, NY, USA, 2002. ACM.

[27] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, pages 497–506, New York, NY, USA, 2009. ACM.
[28] Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schölkopf. Structure and dynamics of information pathways in online media. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 23–32, New York, NY, USA, 2013. ACM.
[29] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. Information diffusion in online social networks: A survey. *SIGMOD Rec.*, 42(2):17–28, July 2013.
[30] Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. From tree house to barracks: The social life of guilds in world of warcraft. *Games and Culture*, 1:338–361, 2006.
[31] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. The life and death of online gaming communities: A look at guilds in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 839–848, New York, NY, USA, 2007. ACM.
[32] Ah Reum Kang, Juyong Park, and Huy Kang Kim. Loyalty or profit? early evolutionary dynamics of online game groups. In *Proceedings of Annual Workshop on Network and Systems Support for Games*, NetGames '13, pages 2:1–2:6, Piscataway, NJ, USA, 2013. IEEE Press.
[33] Akshay Patil, Juan Liu, and Jie Gao. Predicting group stability in online social networks. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 1021–1030, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
[34] Vladimir Batagelj and Andrej Mrvar. Pajek-program for large network analysis. *Connections*, 21(2):47–57, 1998.
[35] J. Nieminen. On the centrality in a graph. *Scandinavian Journal of Psychology*, 15:332–336, 1974.
[36] L. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, pages 35–41, 1977.
[37] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.
[38] Paul Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272, 1901.
[39] Lada A. Adamic and Eytan Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, July 2003.
[40] Caliński, Tadeusz and JA, Harabasz. A Dendrite Method for Cluster Analysis. Communications in Statistics, 3:1, 1-27, DOI: 10.1080/03610927408827101, Jan 1974.