Social Networks 55 (2018) 149–159

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Player-centric networks in *League of Legends*

Marçal Mora-Cantallops*, Miguel-Ángel Sicilia

University of Alcalá, Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Keywords:
Egonets
League of Legends
Player
Video game
MOBA
Affinity propagation

ABSTRACT

Online competitive gaming has become one of the largest collective human activities globally and understanding motivations and social interaction is still not fully achieved. The aim of this study is to develop a basis for a systematic classification of player-centric networks in competitive online games based on structural network criteria. Using data extracted from *League of Legends* players, their matches and machine learning techniques, a classification of personal player networks in *League of Legends* is proposed. Results show the resulting egonets can be potentially grouped in four clusters related to their egos playing habits, ranging from solo to team players.

1. Introduction

Quantification of human behaviour and social dynamics has been a long-lasting challenge for social sciences, hindered by two main factors (Szell and Thurner, 2010): first, dynamics of societies constitute a complex system, characterized by strong and long-range interactions (not treatable, in general, by traditional mathematical methods) and, second, data is of comparably poor availability and quality (Lazer et al., 2009; Watts, 2007).

Both factors are, however, played down when looking at massive multiplayer online games (MMOGs) (Castronova, 2005). In the age of Web 2.0 and, more recently, the era of big data (Chen and Storey, 2012), a great deal of social and relational data is routinely generated and recorded in the course of everyday life. This is the world that Thrift labelled as the world of 'knowing capitalism': a world inundated with complex processes of social and cultural digitalization, with generation, mobilization and analysis of social data becoming ubiquitous (Thrift, 2005). It is also a world where sociologists need to rethink their methodological practices in radically innovative ways, as many assumptions that were central in the 1960s and 1970s no longer pertain in the early years of the 21st century (Savage and Burrows, 2009). These changes go even further, as this digitization reworks the very meaning of social relations, as emphasized by Bruno Latour (2007).

This is especially true in online competitive gaming environments, where a wide range of predefined actions for supporting social interaction reflects either positive or negative connotations among game players, and they are in some cases unobtrusively recorded by game servers (Kwak et al., 2015). Often, data is easily available and can be used to study human and player relations as well as behavioural patterns, providing an unprecedented opportunity to observe social interaction on the large scale (Pobiedina et al., 2013). Stenros et al. (2011) distinguished between different kinds of in-game social interaction and reflected how massively multiplayer games were characterized by the formation of both micro and macro communities, complex communication channel hierarchies and diverse degrees of player involvement in social interactions. Taking into account that online gaming has become one of the largest collective human activities globally, we depart from the assumption that "such games provide for both sufficient participation numbers and careful control of experimental conditions, unlike any other social science research technology" (Castronova, 2006). Castronova contrasts this unique chance to replicate entire societies to the small-scale experiments that are often extrapolated to whole populations and communities. When studying MMOGs, the number of subjects can reach over several hundred thousands and their related actions can be counted by millions. The measurement process is also benefited by how information is extracted; players are not consciously of participating in a research-oriented data gathering process, thus minimizing bias.

As video games evolve and MMOG's popularity grows, video game and player culture also grow, but do so supported by the relationships that arise from their social activity (both *online* and *offline*) (Adamus, 2012). Connection is not only a constitutive fact of social life, but also the pillar where online gaming stands. Players influence each other by means of competition or collaboration, exchange experiences and, sometimes, become involved in longer and meaningful relationships, forming teams or communities. Data extracted from online competitive games such as *League of Legends* can help understanding online players and their habits by looking at the structure of their connections and networks during online play.

As finding stable typologies of player networks based on structural

* Corresponding author.
E-mail address: marcal.mora@edu.uah.es (M. Mora-Cantallops).

https://doi.org/10.1016/j.socnet.2018.06.002

Available online 27 June 2018
0378-8733/

criteria becomes critical when little information other than the network is provided by data generated by online activities, this paper aims to develop a basis for a systematic classification of player-centric networks in competitive online games based on structural network indicators. In particular, every player (ego) will be related to all the teammates he or she had over a year, obtaining a player-centric egonet where two alters are connected by an edge if they played together in at least one match. The resulting egonets can then be analysed according to their structural indicators and categorized (if such a division exists).

Thus, the objectives of this study are:

1 To discover the hidden structure behind the resulting indicators, if any.
2 To define a limited number of indicators that characterize the resulting structure.

To reach them, the following steps will be followed. First, in Section 2, background will be presented, reviewing previous research and providing further detail about what *League of Legends* is and how it is played. In Section 3, participants, indicators and methods applied to extract and build the player-centric networks will be explained. The resulting network dataset will contain the relationships among hundreds of thousands of players so, in order to infer whether a hidden structure from the resulting indicators exists, a machine learning clustering algorithm will be applied in Section 4. This algorithm will be finally optimized (reducing its complexity) until a satisfactory classification of player-centric networks is achieved. The resulting segmentation is then going to be discussed and illustrated. In the final section, the implications of the results will be summarized, practical applications deduced and limitations and future work acknowledged.

2. Background

2.1. Related work

The most explored MMOGs among researchers are in the category of Massively Multiplayer Role-Playing Games (or MMORPGs). Games such as *World or Warcraft* (WoW) can be linked with the much older MUD (Multi-User Dungeon) text-based games, as they fill a similar niche in the gaming world and, at least to some players, provide a fully social experience (Mortensen, 2006). Zhong (2011) examined the impact of collective MMORPG play on gamers' social capital in both the virtual world and the real world. Ang and Zaphiris (2010) used WoW to investigate the social roles that emerged from the users' behaviour and interaction within its guilds (roughly equivalent to in-game clubs) from an analytical perspective and found that the core members of this communities were highly social-oriented players. In spite of this, Ducheneaut et al. (2006) showed that while MMOGs were clearly social environments, joint activities were not as prevalent as they expected. In particular, social network degree densities for in-game guilds were surprisingly low, forming "sparsely knit networks." Other popular games explored include, for example, *EverQuest* (Castronova, 2006) or *Pardus* (Szell and Thurner, 2010).

In spite of the emergence of studies focused on MMORPG in the last decade, few studies have approached massive multiplayer online games from other genres or subgenres such as MOBA (Multiplayer Online Battle Arena) games. Despite its vast enthusiast community and influence on contemporary game designers, remain under-explored by academics, as existing studies acknowledge (Ferrari, 2013). But few games exhibit a greater need for socially-aware services than this relatively new genre (Iosup et al., 2014), as it brings new ways of collaboration and competition on the table, gender and cultural challenges and even new social networks which need to deal with the inherent toxic behaviour that arises in these contexts. MOBA games such as *League of Legends* provide the same opportunity as other MMOGs: namely the scale (*League of Legends* is one of the most played online games globally), data (which is recorded in its servers and accessible using an API) and relevancy (McDonald, 2017). *e-Sports* are a related phenomenon. Taylor (2012) conducted extensive ethnographic research in this regard, while Trepte et al. (2012) used an e-Sport portal to recruit online participants for their work on how offline factors impact online social capital, thus recognizing the relevance of online gaming for research, now that "online gaming has become a major leisure time activity". Carrillo Vera (2015) claims that the impact achieved by *League of Legends* calls for academic and scientific analysis from a range of disciplines, including sociology, economy or communication; taking into account the amounts of data generated every day, however, computer science should also play an important role. This consideration is echoed by Mora-Cantallops and Sicilia (2018), who identified a research opportunity behind MOBA games as a whole, while calling for "future research to include innovative approaches that combine the traditional and common surveys and interviews with data and computer science techniques."

2.2. The game

League of Legends is a multiplayer online battle arena game that follows a freemium model, but where in-game transactions do little to impact a player's performance or ability. In essence, MOBA games are a subgenre of real-time strategy games in which two teams, typically consisting of five players each, compete against each other with each player controlling a single character. Contrary to real-time strategy games, there is no unit or building construction in a MOBA game, so much of the strategy revolves around individual character development and cooperative team play in combat (Yang et al., 2014).

In every *League of Legends* match, two teams face each other in a single map (the *Summoner's Rift*) with a clear goal: to destroy the opposing base (called nexus). As in most MOBA games, teams are composed by 5 players that interact during the game with the aim of optimizing resources, taking advantage of the opponent's errors and destroying other objectives such as towers (that protect the path to the nexus) and neutral monsters (that provide players with rewards).

Although each team in *League of Legends* is composed by five human players, these can be joined in multiple different combinations, from "solo" (which means that the player enters the queue alone and the matchmaking system finds the rest of the team to play with) to a full team composition. Furthermore, each player takes a role in the team. Current matchmaking system allows players to express their preferences and assigns them to a role, which also has an effect on the range of avatar characters (known as Champions in the game) that the player will choose, as some are better suited for a role than others depending on the meta-game at the time (Donaldson, 2015). Role definitions have evolved from season to season, but stabilized at five main roles. Three players control the lanes (Top, Mid and Bottom) while Support provides utility to the team (spending most of the game paired with Bottom) and Jungle makes use of the resources in-between lanes (see Fig. 1). Players can also choose to "Fill", which means that they will take any free role. *League of Legends* is a team game; all five roles are relevant for the team's success. Even though cooperation is critical for success, communication tools are rather limited: natively, the game only allows text chat and pings (pre-set simple sounds or symbols that can be used as a rough system of communication). Thus, some players that join with friends opt to use external conference tools such as Skype to communicate using voice chat.

In *League of Legends*, players are ranked accordingly to their skill level. There are seven tiers in the so called "ladder", in increasing order of skill: Bronze, Silver, Gold, Platinum, Diamond, Master and Challenger. After a few placement matches, players get placed in competition categories (League tiers), and subcategorized into Divisions. The main objective becomes then to climb the ladder by continuously winning matches. Rank distribution changes over the season and can be different depending on the region, but in general,

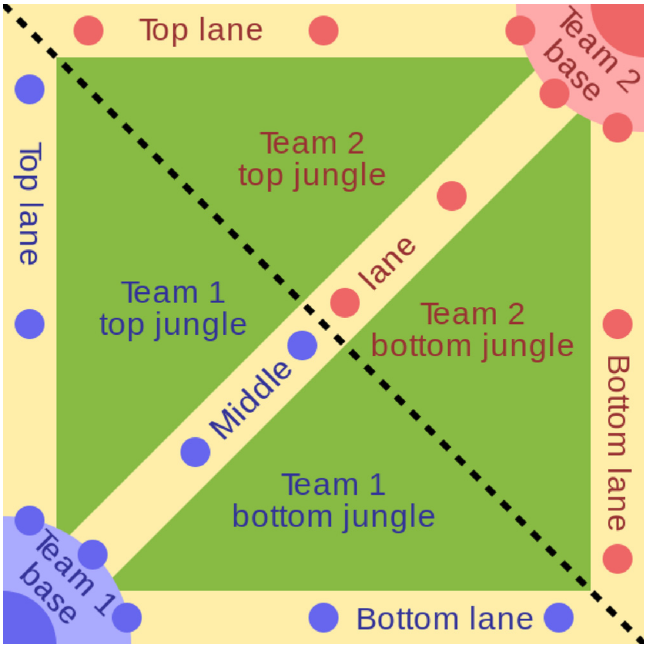

Fig. 1. Typical MOBA map (with labelled lanes) for illustrative purposes. Original PNG version by Raizin, SVG rework by Sameboat. (file:Map of MOBA.png (CC 3.0), CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=29443207).

Silver is the most populated division followed by Bronze and Gold. Altogether, these divisions include between 70% and 90% of the player base. Less than 0.1% players are in Master or Challenger; this last one, actually, is limited to 200 players. Behind this ranking (and using an undisclosed calculation in the case of League of Legends) there is an "Elo" rating system similar to the one originally used for chess players. In short, it is assumed that a player's performance has a normal variation among games; the mean of that distribution is the Elo rating, which is determined by the win/loss statistics. Therefore, a player with a high Elo performs, in average, better than a player with a lower Elo, and this measure is used to match players of similar skill together.

For the current article's purpose it is also relevant to note how *League of Legends* developer Riot Games provides players with free access to the API, a set of tools that can be used to extract player and game data for further research. The available and useful data for this study's purposes, however, is limited to information about their matches and rank.

3. Materials and methods

3.1. Participants

Between July and September 2016, five hundred and forty-seven participants completed an online survey that was distributed (via email) through 20.000 active ranked users in the LVP ("Liga de Videojuegos Profesional") database. The LVP database contains players that are registered in their website and, at the time of the study, it had 68.410 ranked players in it. With over 250,000 registered players, the LVP - Spanish Pro League is the biggest eSports organization in Spain, leading the industry with both online and live tournaments. Manages the most prestigious competitions (Division of Honour), tournaments and other amateur competition systems (LVP Arena) and also broadcasts international events in Spanish such as the League of Legends Championship Series and Call of Duty World League. The LVP also covers gaming technology services, events production, online advertising and audio-visual production covering all aspects of the e-sports ecosystem.

The survey was originally intended to extract information about player habits outside the game (such as media consumption and learning habits), so the questions were irrelevant to the purpose of this study. Summoner name, however, is personal information; thus, participants were asked for agreement on using their player ID to extract information about their match history, which was intended for the current work.

The study targeted ranked players specifically because of two main reasons. First, players achieve ranked status only after they get past the thirty level game tutorial. Therefore, there is no "novice" effect that could have an impact on results. Second, the League of Legends API only fully records data for ranked matches, converting unranked data in technically unreachable. As Summoner name is considered personal information, it was entered optionally and manually in the questionnaire, reducing the final number of complete entries to four hundred and thirty-nine. Final demographics were, therefore, N = 439, age between 13 and 35 years of age (average at 19.4 with a standard deviation of 3.45). 93.8% were male (N = 411, average 19.2 years of age, SD = 3.36) and 6.2% were female (N = 27, average 21.96 years of age, SD = 3.87).

3.2. Network generation

All matches played in 2016 by each of the 439 respondents (referred as egos from this point) were extracted through the *League of Legends* API[1] provided by Riot. As a result, a total of 228.117 matches were obtained, with a mean of 520 matches per ego (SD = 424).

When a player joins a match, he or she does so in a match lobby, where the player is joined by other players until a team of five components is formed to play against five other players. Therefore, for each ego and for each match, the relationship between all team members is registered. Every relationship is counted as many times as it appears; thus, the weight of the link reflects how many matches two players have played together. Weight will be required for the modularity algorithm in Section 3.3.5 and will only be used for graphical representation in Fig. 9.

After processing all egos, the average number of alters per network is 1535 alters, for a total of 674.205 nodes (egos and alters – but players in the end) overall, but with considerable differences: the smallest network has 18 nodes while the largest has 7896. Approximately 80% of the networks have a number of nodes between 200 and 3500, however. Due to the described construction, all nodes are connected through the ego. Thus, before the subsequent analysis, the ego is removed from the network, highlighting the underlying alter to alter structure under the ego effect.

In summary, the resulting networks are one-mode projections of the two-mode networks connecting players to matches. Matches have, however, one restriction: its degree in the bipartite network is always equal to four. As a result, after removal of the ego, the one-mode network is a network of overlapping four-cliques or K4 complete graphs.

For reference, as shown in Fig. 2a, if an ego played a single match, the network would be a K5 complete graph (a five-node graph in which every pair of vertices would be connected). If an ego played two matches with the same team, the network would still be the same (with double weight in its links), but if an ego played the second match with a complete different team, then the resulting network would look like two K5 graphs linked by a bridge – the ego (Fig. 2c). Removing the ego in the first case would keep the network connected (Fig. 2b); doing the same in the second case would leave two disconnected components (Fig. 2d). The generalization of this example will become key to understand the indicators that follow and their impact in the player networks.

[1] Application Programming Interface. More League of Legends API information is available at https://developer.riotgames.com/.

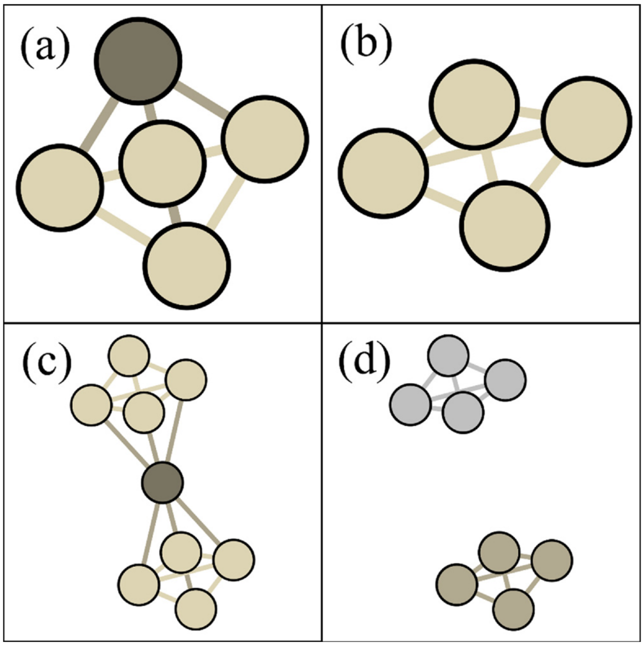

Fig. 2. Network generation. (a) Ego with a single team (b) Alter network from (a) after ego removal (c) Ego after two matches with complete different players (d) Alter network from (c) after ego removal.

3.3. Indicators

Mathematically, networks are described by graphs (Wasserman and Faust, 1994). An undirected graph G is described by a set of nodes $N = \{n_1, n_2, \ldots, n_g\}$ and a set of links (also called edges) $L = \{l_1, l_2, \ldots, l_L\}$ between pairs of nodes, where $l_k = (n_i, nj)$. A large number of structural indicators can be computed on a network, some with a clear meaning and others with a more technical one, but in any case all features have an implication when used in the context of the social network analysis.

As *League of Legends* is a team-based game with a relevant social component, for the current analysis purpose it was assumed that the most relevant structural features would be those related to node relevancy (as it was expected to find "friends" as relevant nodes) and to cohesion (as social players would see highly knitted networks as opposed to disperse components in non-social users). Node relevancy will be measured through degree and betweenness, while node density, component measures and modularity will represent structural network cohesion.

3.3.1. Degree

The degree of a node, denoted by $d(n_i)$, is the number of edges that are incident with it. Equivalently, it's the number of nodes that are adjacent to it. Degrees are easy to compute and informative; alters with a small degree will indicate players that played with few other of the ego's alters, while a high degree will show the opposite. In this case, therefore, degree will become a measure somehow related to the ego's social circles: the higher the degree, the closer to the core of the ego's playing community that alter is. For this application, the mean nodal degree (or average degree) will be used to summarize the degrees of all the actors in the network. For a network with g nodes and L links, mean degree $\bar{d}$ can be calculated as:

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g} \tag{1}$$

3.3.2. Node density

Every processed match can add up to four different players to the ego network (the fifth player is always the ego). If any of them already exists, the number of new nodes will be less than four. Therefore, by construction, the maximum possible number of nodes $g_{\max}$ equals four times the number of matches (m) the player joined. Node density for a player-centric graph with g nodes is then defined as follows:

$$node \quad density = \frac{g}{g_{\max}} = \frac{g}{4^*m} \tag{2}$$

3.3.3. Betweenness centrality

Interactions between two nonadjacent nodes might depend on the nodes that lie on the paths between the two. When this happens, these in-between nodes might have some control over their interactions (Wasserman and Faust, 1994). This becomes especially relevant in the player-centric network, as it's highly possible that the ego reaches new players through his or her frequent colleagues or friends. Betweenness centralization (c_B) for a node $n_k \in G$ can be defined as the number of shortest paths between n_i and n_j that pass through n_k ($\sigma(n_i, n_j|n_k)$) divided by the total number of shortest paths between n_i and n_j ($\sigma(n_i, n_j)$) (Brandes, 2008). Formally:

$$c_B(n_k) = \sum_{n_i, n_j \in G} \frac{\sigma(n_i, n_j|n_k)}{\sigma(n_i, n_j)} \tag{3}$$

By convention, this definition applies to disconnected graphs without modification (Freeman, 1977). Although the distance between two disconnected nodes is undefined, the number of shortest paths between them is defined and equal to zero. The resulting contribution to betweenness centrality is then established as zero. The mean betweenness centrality measure is then the average of c_B across all nodes. Additionally, the number of nodes with c_B above three times the average will be counted in each network as it will give an indication of the number of players with a relevant flow control.

3.3.4. Components and largest connected component

A graph is connected if there is a path between every pair of nodes in the graph. Else, every maximal connected subgraph is a component. Note that if there is only one component the graph is connected. In graphs with infinitely many nodes, the emergence of a giant component is observed after crossing a certain threshold (Dorogovtsev and Mendes, 2003). The same concept can be applied to finite graphs, where the component with the highest number of nodes is called the largest connected component.

For the purpose of this study, an additional measure is calculated. Let g be the total number of nodes in G and g' the number of nodes in the largest connected component. It's then possible to calculate the largest component proportion as g' divided by g. While the number of components might give an indicator of the different groups of play that the ego has, this proportion will provide an indication of how large is his or her main playing group.

3.3.5. Modularity

A key feature of social networks is high transitivity, meaning that if n_i is connected to n_j and n_j is connected to n_k, there is a high chance of having a connection between n_i and n_k too. This property leads to the formation of clusters called *communities*, "with groups of nodes within which connections are dense but between which they are sparser" (Newman, 2003). Multiple community detection algorithms have been described (Blondel et al., 2008; Clauset et al., 2004; Girvan and Newman, 2002; Pons and Latapy, 2006) but the result is always some division of the vertices into communities. The quality of this division is often measured by the modularity of the partition (Newman, 2003), a scalar value between -1 and 1 that measures the density of links inside the obtained communities as compared to the links between them. This quality function Q, modularity, is defined as follows. Let e_{st} be the fraction of edges in the network that connect nodes in group s to those in group t, and let $a_s = \sum_t e_{st}$. Then:

$$Q = \sum_{s} (e_{ss} - a_s^2) \tag{4}$$

is "the fraction of edges that fall within communities, minus the expected value of the same quantity if edges fall at random without regard for the community structure" (Newman, 2003). Note that the expected modularity for a random partition would be 0 and any other value reflects a deviation from pure chance. According to Newman, values greater than 0.3 appear to indicate relevant community structure. In the current study, the Louvain method for community detection will be used as defined by Blondel et al. (2008) and implemented using the NetworkX 2.1 (Python 3.6) libraries and the Gephi tool. The Louvain method is an efficient community detection algorithm broadly used that features a modification on (6) (in order to consider weights) as the function to optimize.

3.4. Cluster analysis

Cluster analysis is a category of unsupervised machine learning techniques that allow to discover hidden structures in data where the ground truth is unavailable (so, where the right answer, if any, is unknown) such as the one in question. The goal of this technique is, therefore, "to find a natural grouping in data such that elements in the same cluster are more similar to each other than those from different clusters" (Raschka, 2014).

Many clustering algorithms exist. The standard Python *scikit-learn* library has implemented the most popular and it was the package used in this analysis. One of the most used methods is the K-means algorithm (Arthur and Vassilvitskii, 2007), that clusters data by trying to separate samples in n groups of equal variance. For the current sample, however, K-means presented two drawbacks. First, it requires the number of clusters to be specified. Therefore, one should have an idea of how many clusters are expected in the data before applying it, which wasn't the case. Second, due to its implementation, K-means expects a certain normality in the input data, which couldn't be assumed in the player-centric dataset. K-means is also unstable, and clustering depends on initialization, which was undesirable.

The affinity propagation algorithm (Frey and Dueck, 2007) is a newer method that has some advantages over K-means: the number of clusters doesn't need to be specified beforehand, non-symmetric dissimilarities are supported and it is stable over runs. The affinity propagation algorithm identifies exemplars among data points and forms clusters around these exemplars. It operates by simultaneously considering all points as potential exemplars and exchanging messages between them until a good set of exemplars and clusters emerges. As its characteristics are more appropriate to classify the player-centric network dataset, affinity propagation is going to be the method used for clustering.

Still, two parameters need to be set in advance. *Damping* is set in all calculations to 0.8 in order to avoid undesired oscillations while computing. *Preference* is defined as the suitability of a particular data point to serve as an exemplar. High preference values will result in many exemplars found (many clusters), while lower values will lead to a small number of exemplars. When preferences rise above a certain value, it becomes beneficial for multiple subsets of data that have approximately the same intra-subset similarities and approximately the same inter-subset similarities to form distinct clusters simultaneously, so the number of clusters obtained quickly rises. Thus, "different plateaus would correspond to the extraction of different levels of structure" (Frey and Dueck, 2007). Therefore, preference value will be need to be analysed and chosen in each particular instance of the analysis, looking for these "plateaus" in Figs. 4 and 6.

A whole different branch of SNA is devoted to blockmodelling, another alternative for analysis. In essence, blockmodelling compares patterns of connection between nodes to cluster them into "blocks" of nodes that enjoy similar position or roles within the network. The goal of blockmodeling is to reduce a large, potentially incoherent network to a smaller comprehensible structure that can be interpreted more readily. In spite of this, blockmodelling techniques are "very unusual in ego-net analysis because ego-nets are generally too small to merit blockmodelling" (Crossley et al., 2015). There are exceptions, however, as ego-nets might be big enough to merit its use, as in the work by Edwards and Crossley (2009).

4. Results and discussion

4.1. Variables

For each player-centric network, the following properties were calculated after removal of the ego: average degree, node density, average betweenness centrality, percentage of nodes with betweenness centrality over three times the mean, number of separate components, largest component proportion and modularity. As the number of separate components is affected by the number of matches played, it was divided by the total number of matches for every particular player. Once this was done, all variables were standardized by removing the mean and scaling to unit variance. The resulting dataset contained 439 observations with seven indicators per row.

To address the first objective, a clustering algorithm (affiliation propagation) will be used. Although there might be alternatives to using clustering techniques, they allow to discover hidden structures in data where the ground truth is unknown and are thus appropriate for this duty.

4.2. Bivariate correlation

Before proceeding to clustering, the bivariate correlation table between indicators or variables of interest is presented in Fig. 3 as reference. Two groups can be distinguished; the first one containing measures related to node cohesion and the second one related to node relevancy. All variables will be included in the initial clustering model, which will be optimized later to reduce the number of indicators.

4.3. Clustering

An affiliation propagation model was build using the obtained dataset. Before proceeding, however, the preference parameter had to be set. To do so, the influence of this parameter to the number of clusters obtained by the model was plotted in order to find a significant plateau (Fig. 4).

A long plateau between $[-60, -45]$ can be observed, resulting in a seven cluster structure. Affinity propagation is then used with *preference* equal to -45.

As displayed in Fig. 5, seven clusters are formed. To assess the goodness of fit, the average silhouette width (ASW) (Kaufman and Rousseeuw, 1990) is computed and results in $ASW = 0.533$. ASW assesses the optimal ratio of the intra-cluster dissimilarity of the objects within their clusters and the dissimilarity between elements of objects between clusters. According to Kaufman and Rousseeuw, an ASW between 0.51 and 0.7 indicates that "a reasonable structure has been found," so, with 0.533, the present clustering shows a reasonable preliminary classification of player-centric networks.

4.4. Optimizing indicators

The second objective was to reduce the number of indicators that defined the hidden structure that emerged from the networks. A series of ordered cuts in the variables was executed, iterating as follows: remove indicator, assess preference parameter, run affinity propagation, compare ASW. The final and best result was obtained using only the modularity, the standardized number of components and the largest component proportion, thus reducing the number of indicators from

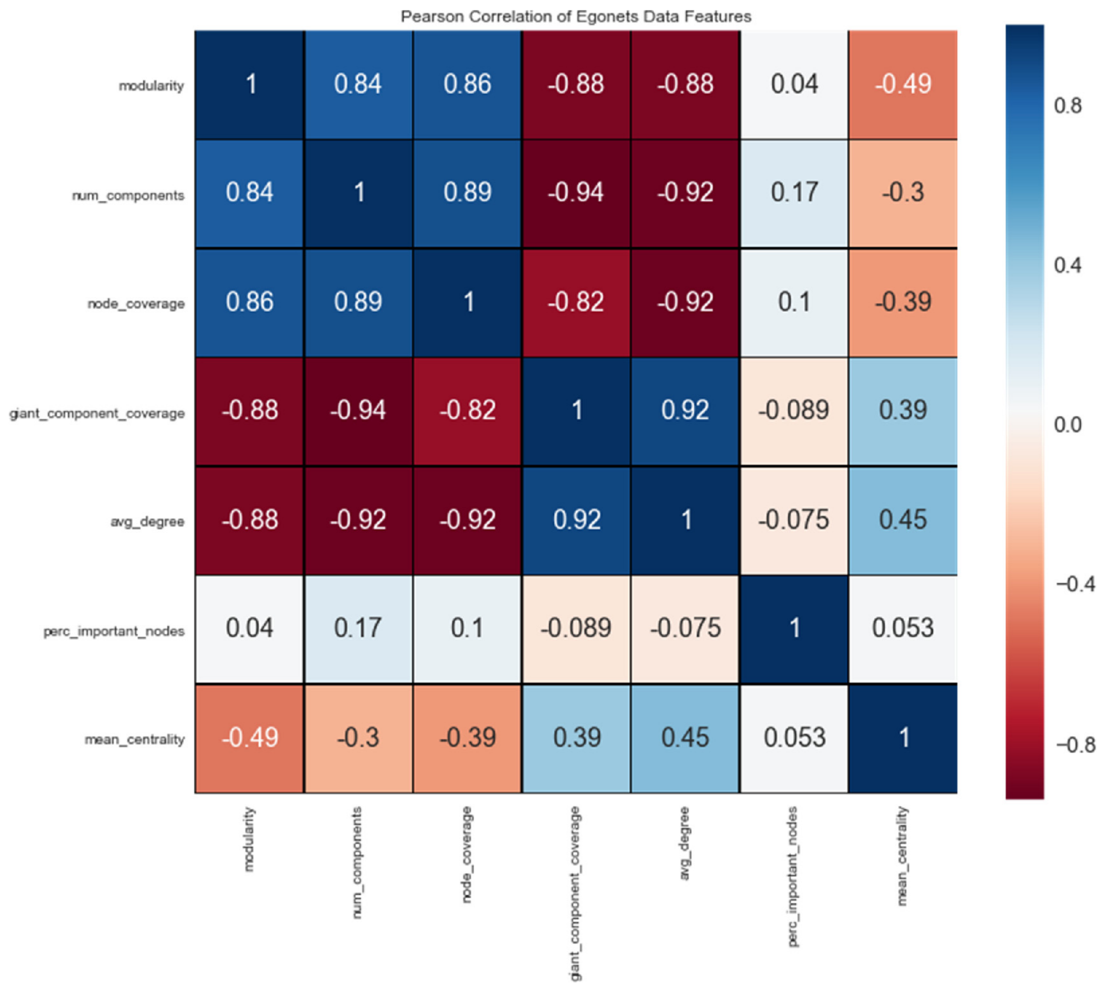

Fig. 3. Bivariate correlations between variables of interest.

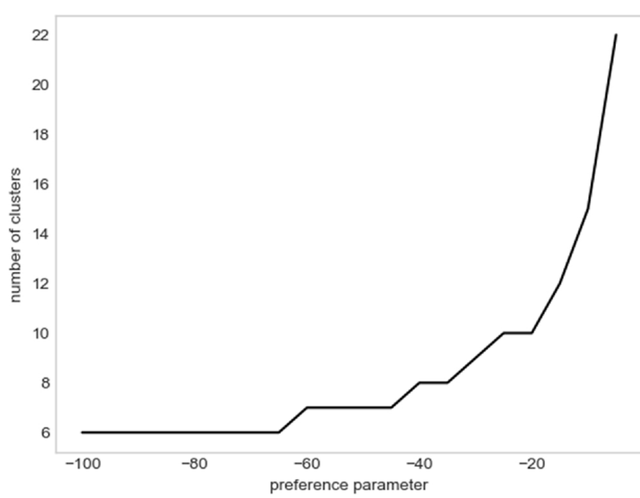

Fig. 4. Number of clusters derived from affiliation propagation versus shared preference (all indicators included).

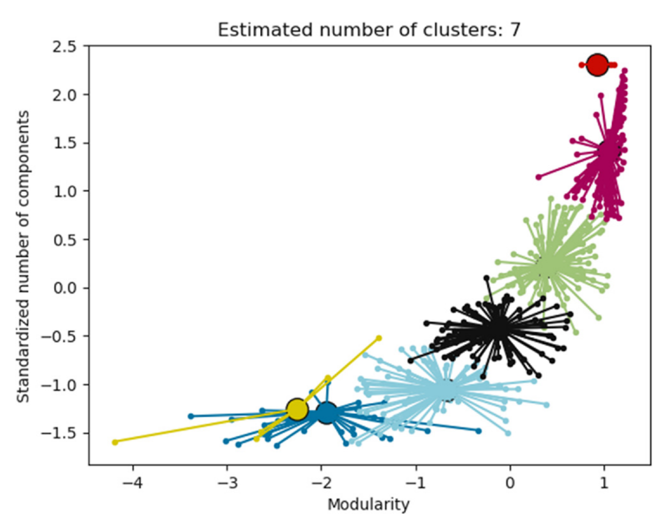

Fig. 5. Graphical representation of clusters resulting from the affiliation propagation algorithm with preference −45 and all indicators included. This is only a 2D slice of a 7 dimension space for illustration purposes.

seven to three.

With *preference* set at −30 (see Fig. 6, although using the same value as in the previous run wouldn't change the result), the affiation propagation algorithm results in four clearly defined clusters (Figs. 7 and Figure 8) that assimilate the three small additional clusters that appeared in Fig. 5 into them. The ASW also improved notably and equals 0.641, still in the same *reasonable* structure interval but with a simpler cluster split and closer to the 0.71 that would be the threshold to obtain an excellent fit. From largest to smallest, the number of observations per cluster is 65 (C1), 125 (C2), 129 (C3) and 120 (C4).

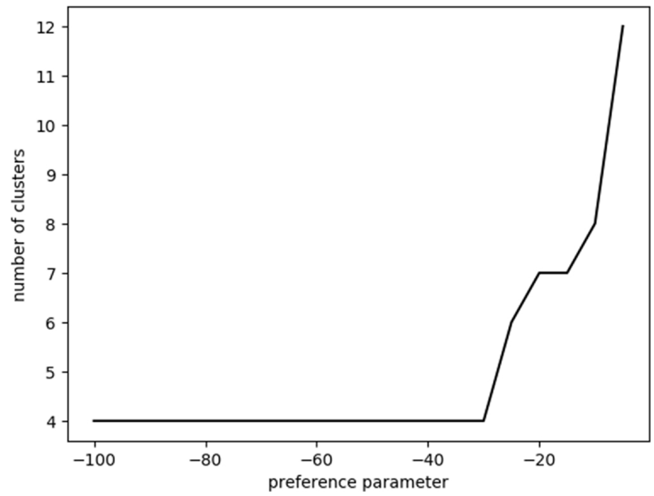

Fig. 6. Number of clusters derived from affiliation propagation versus shared preference (best three indicators only).

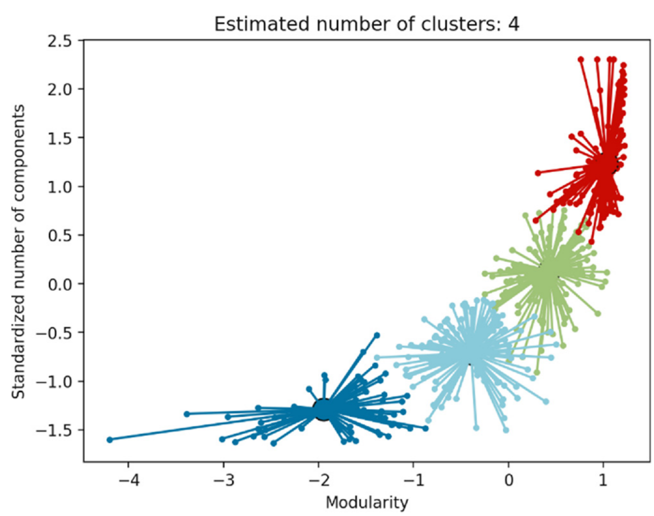

Fig. 7. Graphical representation of clusters resulting from the affiliation propagation algorithm with preference −30 and restricted to the best three indicators. This is a 2D slice of the 3D space.

4.5. Network attributes by cluster

Another way to check whether the final proposed clustering has a good fit is to look at the distribution of the indicators per each cluster and compare them. To do so, four violin plots are drawn in Fig. 9.

Before anything else, a quick test is run on the total number of games per player; the graph shows that they are similar among clusters so size-related side-effects can be discarded. A Kruskal-Wallis test followed by a Dunn's test is then run for all three indicators. Differences are found for all three at p-value ~ 0, and all pairs of clusters present statistically significant differences with p-values $\ll 0.001$.

Cluster C1 contains player-centric networks that have low modularity (so few communities emerge), a low number of components in proportion to their size and the largest component contains, in average, more than 80% of the total nodes.

Cluster C4 networks have high modularity (close to the highest possible value), a massive number of components in proportion to their size and the largest component contains, in average, less than 20% of the total nodes.

In-between these two extremes, clusters C2 and C3 are found. Both have higher modularity and component counts than C1 but notably lower than C4. Moreover, both have largest components with less nodes than C1 (around 60% and 40%, respectively) but with a significant increase versus C4. Although these similarities, they are still quite different, and it is assumed that they represent two distinct groups of players.

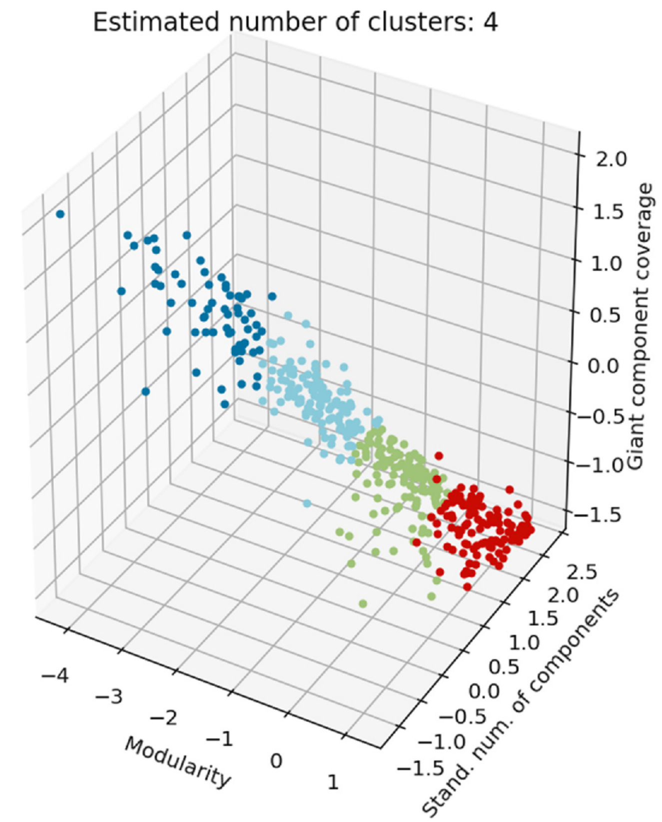

Fig. 8. 3D representation of the resulting clusters. Note that Fig. 7 is the 2D view from above.

An observation is chosen at random from each cluster to illustrate the described typology. For both visualization and comparison purposes, it's selected randomly from similar sized networks (range between 500 and 800 games played). Their dimensions and indicators can be found in Table 1, while the graphical representation is presented in Fig. 10, where thicker edges represent higher weights.

- C1 would correspond to a "team player". The alter network (without the ego) is not only highly connected and with a huge largest component, but also exhibits strong connections between some alters. This implies that the ego a) almost always plays with the same players and b) they do so together. As many matches are played sharing alters, the total network contains notably less nodes than the other clusters (as can be noted in Fig. 9). These players often join the game with a team full of known people (or, at least, players with whom they already played in the past).
- C2 would then correspond to a "group player" instead. Compared to C1, there are fewer strong links between alters (in this particular case it's basically a triangle), the largest component is smaller and there are more disconnected components. This kind of player a) regularly plays with two or three friends and b) occasionally plays with people in other circles or alone. Therefore, "group players" (whenever possible) join with a group that is not enough to cover a full team. Else, they play with smaller groups or even solo as a last option.
- C3 shows much less strong relationships between alters, so this cluster could be labelled as "cell players". The largest component covers less than half of the network and connections inside are weaker. Two tendencies are found: strong dyadic relationships in

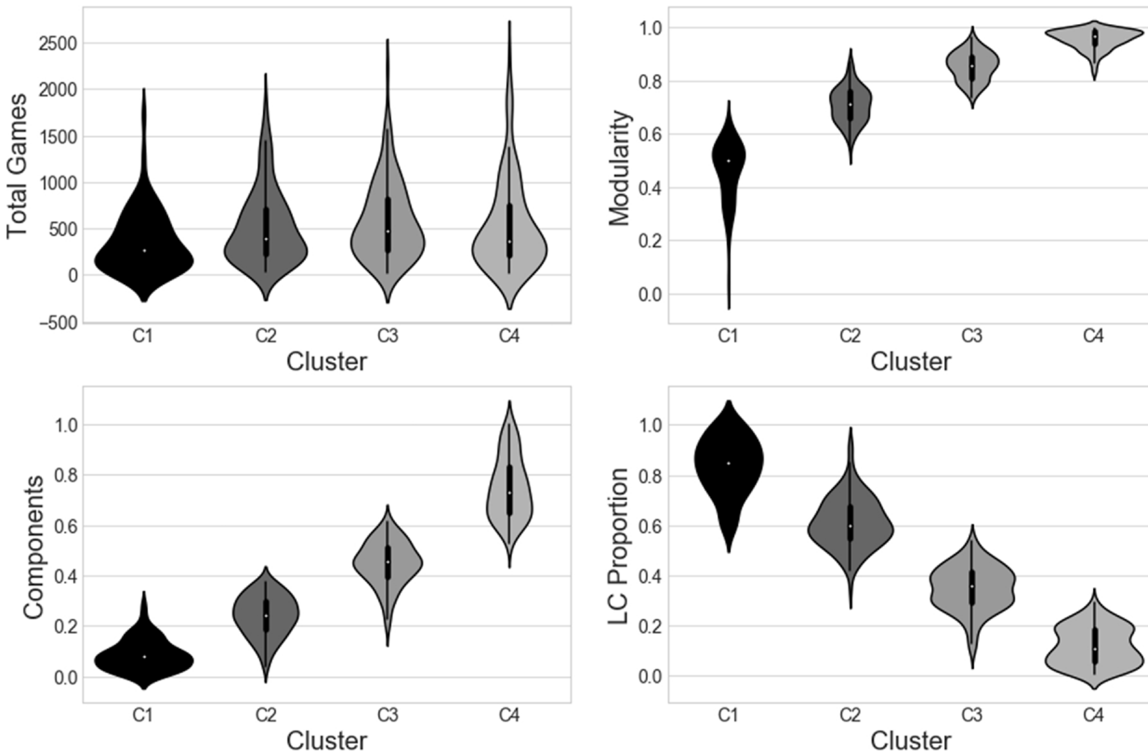

Fig. 9. Violin plots for total games, modularity, components and largest component proportion per cluster.

Table 1
Indicators extracted from sample player-centric networks.

	Matches	Nodes	Modularity	Standardized Number of Components	Largest Component Proportion
C1	693	1335	0.323	0.059	0.878
C2	668	1606	0.598	0.178	0.690
C3	699	2069	0.775	0.383	0.423
C4	699	2641	0.992	0.816	0.058

the alter network are translated into matches played in "trios" (but note that these trios are not fully connected between them; if they were, they would become C2). Additionally, many small star-shaped components outside the largest one represent games played in "duo" (therefore sharing only one player between matches). Players in C3 join the game in small and variable cells of two to three members; they also go "solo" more often than the previous two types.

- C4 would then be the "solo" player. Their graphs show how the largest component is almost inexistent, no relevantly strong links are present with any other player and the landscape is mostly composed of spare components of four unknown players that the ego will never met again. This is the least social of all players. Therefore, not only their games are always automatically filled with four strangers by the matchmaking system, but also rarely results in links established between them.

4.6. Clusters and player rank

Although the official League of Legends API provides little information about player's attributes, rank is available. Therefore, it is possible to enhance the obtained clustering with a few insights about their skill level (as determined by the game). Fig. 11 displays rank distribution within each cluster.

Taking into account that bronze, silver and gold are the lowest ranks and platinum and diamond are the highest ones before the exclusive master and challenger (none in the sample):

- The first cluster, C1, formed by team players, presents notably less high ranked players than the other clusters.
- On the other hand, C3 and C4, the duo and solo players, have a higher proportion of platinum plus diamond players, even over the expected distribution across all player base.

At first sight, this is somehow contradictory, as collaboration and trust between team members in *League of Legends* is crucial for success so it would make sense to see better collaboration (and, therefore, better performance) in teams formed by friends. In spite of this, two reasons why this happens could be suggested.

First, the higher the rank the less players there are. When a player climbs in the ladder, not only there are less players available, but less friends too, as players need to join the games with players either in their own division or one up/down. This keeps true until platinum, where restrictions apply, as only highly ranked platinum players can join games with diamond players. Diamond gets more extreme, as there are even restrictions within its divisions. Therefore, higher ranked players are forced to be solitary players in regard to their friendships.

Second, in ranked competition all games matter. It is not only important to play well, but also to keep winning in order to advance. A bad streak can bring a player down in the ladder and promotions are long and difficult. It is thus likely that friendly team playing is left for lower categories (suggesting more "casual" playing) while higher ranked players might tend to select their teammates more in detail. It is even possible that, in order to prevent disputes and negative feelings, joining with real-life friends is avoided in these almost professional ranks of skill.

5. Conclusions

Typologies are useful to compare networks systematically and player-centric networks become more and more relevant as online gaming grows. Although data in games such as *League of Legends* is recorded by game servers, demographic data is not available, so this article has limited its scope to the purely structural characteristics of

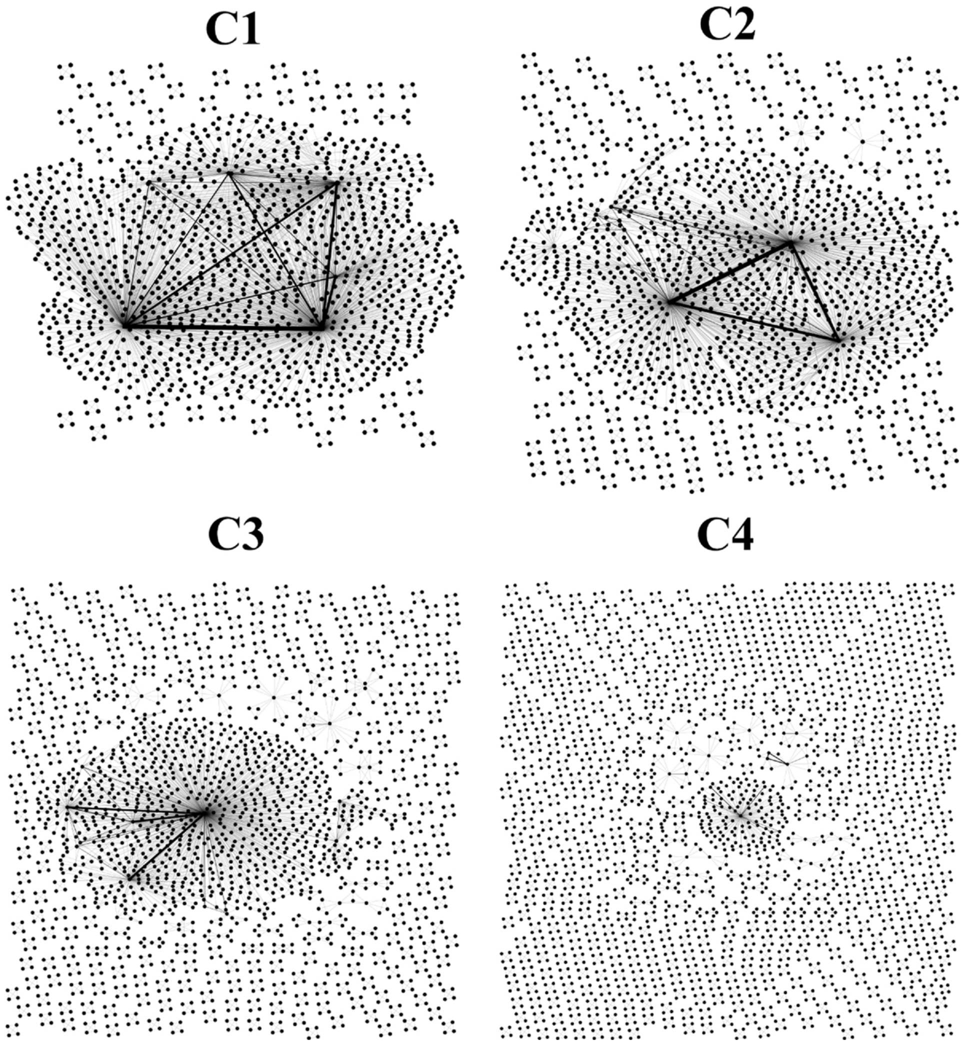

Fig. 10. Graphical representation of sample player-centric networks.

the ego-networks formed by 439 players (439 egos and 673.766 alters) and all their matches (228.117) over a period of a season (all matches in 2016).

First of all, player data was extracted using the API provided by Riot (*League of Legends* developer) and networks were built in Python using NetworkX. The resulting networks were then analysed and eight structural indicators were computed for each of them. In order to infer the underlying hidden structure, machine learning techniques were used, applying an affinity propagation algorithm. After a few iterations and removal of individual variables, the best clustering was achieved with three indicators.

The player-centric networks obtained in the current study were complex and heterogeneous, but their underlying structure can be described and typified using the following three features: modularity, number of components (standardized) and the proportion of nodes covered by the largest emerging component found in the network. Together, they establish four degrees or categories that describe how social the ego (or the player) is in his or her gaming habits.

The proposed presented typology provides an intuitive and systematic method to characterize the social behaviour of *League of Legends* players looking only at the structure of their player-centric network. Even though only the most popular online competitive game was assessed, this methodology can be generalized to any online competitive game that provides enough data to compute its player-centric networks.

The obtained information can then be used for player segmentation, both to improve player experience (by adapting the game to their structural social needs) and to improve the game (adapting its matchmaking system). Players that go "solo" are focused in the game; they need quick access to their matches so team formation or discussion might be less important for them, but they would expect to be paired against other "solo" or isolated players. On the other side, pure "team players" could need access to additional social features to empower their social relationships, to better means of communication (that are natively limited to text chat in *League of Legends*) and to fair matchmaking against other "team players" instead of spare groups of "solo" players or "cell players".

This typology, however, could be influenced by ego or alter attributes. Rank, for example, seems to have an influence on the structural characteristics of the resulting player-centric networks. Higher ranked players are more often "duo" or "solo" players, while "team players" are overrepresented in the lower categories. This suggests that other attributes that were not captured in our study (or were not available for extraction) could also present a relevant influence, so further research is required.

Implications of these findings also go beyond the scope of the game. For the industry it is not only relevant to dimension their players' egocentric networks but also to be able to find patterns that cluster them together. MOBA games such as *League of Legends* are always played against the developer's servers, so this profiling could then be used to improve service, clustering players that show either similar or complementary patterns together. Loyalty rewards could also be adapted according to player's behaviour. As of today, players receive

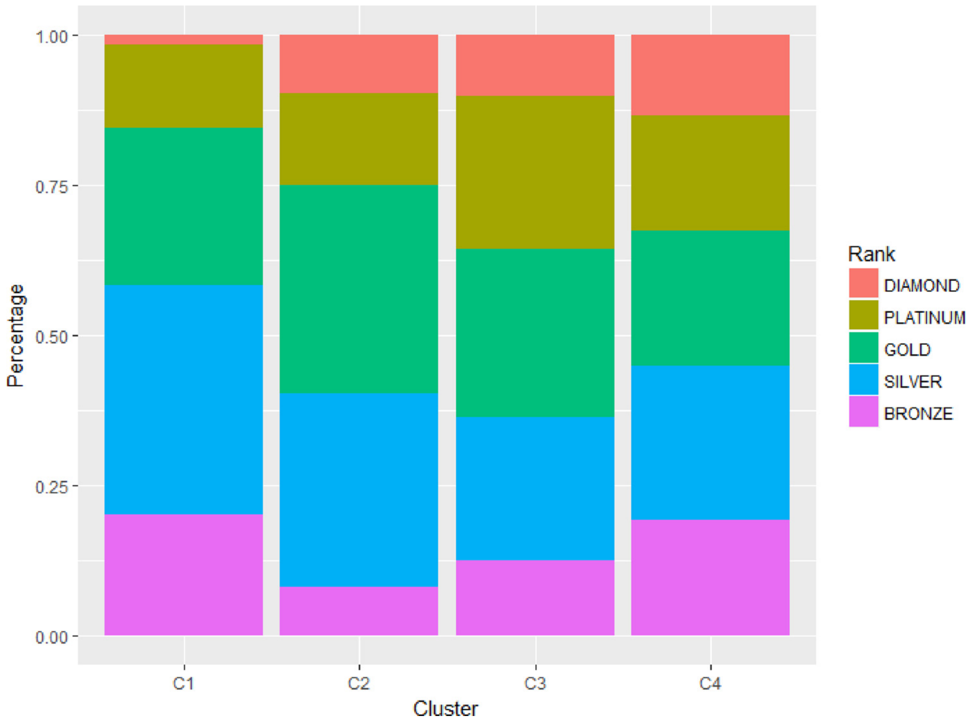

Fig. 11. Rank distribution per cluster. C1 has less high ranked players than other clusters. C3 and C4 are disproportionally higher ranked.

rewards individually; this might be enough for solo players, but team players might be more satisfied if rewards were matched across the team (for example, with matching themed *skins* for their characters). Matchmaking could also see a social improvement: if a regular team player (so, a highly social player) joins the game alone for once, it might be more appropriate to match him or her with a four player team which misses one player than with four solo players, as his or her experience will then be similar to the one he or she is used to. This is the kind of socially-aware services that are needed as online gaming becomes more and more popular.

6. Limitations and future work

While results show the potential of retrieving game data to understand player-centric networks and to profile player social behaviour systematically in any online competitive game, additional work would be needed to assess whether the conclusions of this study could be extrapolated to other online MOBA games or other genres. At least initially, one could extrapolate these results to other online competitive games, as long as configuration is similar (teams of five collaborating players facing five other players). The general idea is that, through the player-centric network described by his or her matches, a player can be classified as a more or less social player, which might need additional features in order to keep playing or to improve his or her experience. For example, we noticed that in *League of Legends*, most social players are ranked low. This brings additional questions. Are teams formed by friends more frustrated than solo players? Are they more prone to abandon the game? This study does not have enough data to verify whether this happens, but at least it is something that the developer (or developers of future games) could take into account when designing their rank system. Would it make sense, for example, to have separate ranks or ladders by type of social player? Could rank depend on more than pure results (e.g. playing with honour, being a good teammate, helping others)? Players might have a better experience and feeling of fairness if ranking systems adapted to their preferences and, as a result, they might spend more time in the platform which could translate into additional purchases or loyalty.

Another limitation (due to the method used for extraction of data) is the unavailability of demographic or personal information about the players. Sensitive personal data cannot be obtained from the API so this study has been limited to purely structural network information. Future work should find a way to include individual attributes of the egos or alters (e.g. gender, age, nationality, studies, other games played, etc.) to fine-tune the proposed typology or to link the findings to “real-world” issues, such as the relationships among the player-centric networks and the players offline social circles.

As discussed, online games such as *League of Legends* represent an unprecedented chance and a unique opportunity to study complex social systems on an entirely different scale. The scale is so massive, however, that the study of hidden structures and systematic classification becomes critical for their understanding. If this study has been able to find a non-trivial structure related to the playing habits of each ego, further structures could be found, showing the potential of this method to get closer to a comprehensive understanding of the complex and unscaled social interactions happening online among players at every moment.

References

Adamus, T., 2012. Playing computer games as electronic sport: in search of a theoretical framework for a new research field. Computer Games and New Media Cultures: A Handbook of Digital Games Studies. pp. 477–490. http://dx.doi.org/10.1007/978-94-007-2777-9_30.

Ang, C.S., Zaphiris, P., 2010. Social roles of players in mmorpg guilds: a social network analytic perspective. Inf. Commun. Soc. 13, 592–614. http://dx.doi.org/10.1080/13691180903266952.

Arthur, D., Vassilvitskii, S., 2007. K-Means + +: the advantages of careful seeding. Proc. Eighteenth Annu. ACM-SIAM Symp. Discret. Algorithms, vol. 8, 1027–1035. http://dx.doi.org/10.1145/1283383.1283494.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, 1–12. http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. Soc. Netw. 30, 136–145. http://dx.doi.org/10.1016/j.socnet.2007.11.001.

Carrillo Vera, J.A., 2015. La dimensión social de los videojuegos “online”: de las comunidades de jugadores a los “e-sports.”. Index Comun. 10, 39–51.

Castronova, E., 2005. Synthetic Worlds: The Business and Culture of Online Games. University of Chicago Press.

Castronova, E., 2006. On the research value of large games: natural experiments in Norrath and Camelot. Games Cult. 1, 163–186. http://dx.doi.org/10.1177/1555412006286686.

Chen, H., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. Mis. Q. 36, 1165–1188. http://dx.doi.org/10.1145/2463676.2463712.

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70, 66111. http://dx.doi.org/10.1103/PhysRevE.70.066111.

Crossley, N., Bellotti, E., Edwards, G., Everett, M.G., Koskinen, J., Tranmer, M., 2015. Social Network Analysis for Ego-Nets. SAGE Publications.

Donaldson, S., 2015. Mechanics and metagame: exploring binary expertise in league of legends. Games Cult. 1–19. http://dx.doi.org/10.1177/1555412015590063.

Dorogovtsev, S.N., Mendes, J.F.F., 2003. Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford Univ. Presshttp://dx.doi.org/10.1093/acprof:oso/9780198515906.001.0001.

Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J., 2006. “Alone together?” Exploring the social dynamics of massively multiplayer online games. Chi 2006 407–416. http://dx.doi.org/10.1145/1124772.1124834.

Edwards, G., Crossley, N., 2009. Measures and meanings: exploring the ego-net of Helen Kirkpatrick Watts, militant suffragette. Methodol. Innov. Online 4, 37–61. http://dx.doi.org/10.1177/205979910900400104.

Ferrari, S., 2013. . From generative to conventional play: MOBA and league of legends. DiGRA Conf. Proc. 2013.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35. http://dx.doi.org/10.2307/3033543.

Frey, B.J., Dueck, D., 2007. Clustering by passing messages between data points. Science (80-.) (315), 972–976. http://dx.doi.org/10.1126/science.1136800.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99, 7821–7826. http://dx.doi.org/10.1073/pnas.122653799.

Iosup, A., Van De Bovenkamp, R., Shen, S., Jia, A.L., Kuipers, F., 2014. An Analysis of Implicit Social Networks in Multiplayer Online Games, vol. 18. pp. 1–8. http://dx.doi.org/10.1109/MIC.2014.19.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics). Eepe. Ethz. Ch. http://dx.doi.org/10.2307/2532178.

Kwak, H., Blackburn, J., Han, S., 2015. Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games. http://dx.doi.org/10.1145/2702123.2702529.

Latour, B., 2007. Beware, your imagination leaves digital traces. Times High. Lit. Suppl. 6.

Lazer, D., Brewer, D., Christakis, N., Fowler, J., King, G., 2009. Life in the network: the coming age of computational social science. Science (80-.) 323, 721–723. http://dx.doi.org/10.1126/science.1167742.Life.

McDonald, E., 2017. The Global Games Market 2017 | Per Region & Segment | Newzoo [WWW Document]. Newzoo.com. URL. https://newzoo.com/insights/articles/the-global-games-market-will-reach-108-9-billion-in-2017-with-mobile-taking-42/.

Mora-Cantallops, M., Sicilia, M.-Á., 2018. MOBA games: a literature review. Entertain. Comput. 26, 128–138. http://dx.doi.org/10.1016/J.ENTCOM.2018.02.005.

Mortensen, T.E., 2006. WoW is the new MUD: social gaming from text to video. Games Cult. 1, 397–413. http://dx.doi.org/10.1177/1555412006292622.

Newman, M.E.J., 2003. Fast Algorithm for Detecting Community Structure in Networks. pp. 1–5. http://dx.doi.org/10.1103/PhysRevE.69.066133.

Pobiedina, N., Neidhardt, J., Del Carmen Calatrava Moreno, M., Grad-Gyenge, L., Werthner, H., 2013. On successful team formation: statistical analysis of a multi-player online game. Business Informatics (CBI), 2013 IEEE 15th Conference on. pp. 55–62.

Pons, P., Latapy, M., 2006. Computing communities in large networks using random walks. J. Graph Algorithms Appl. 10, 191–218. http://dx.doi.org/10.7155/jgaa.00124.

Raschka, S., 2014. Python machine learning. Igarss 2014. http://dx.doi.org/10.1007/s13398-014-0173-7.2.

Savage, M., Burrows, R., 2009. Some further reflections on the coming crisis of empirical sociology. Sociology 43, 762–772. http://dx.doi.org/10.1177/0038038509105420.

Stenros, J., Paavilainen, J., Mayra, F., 2011. Social interaction in games. Int. J. Arts Technol. 4, 342. http://dx.doi.org/10.1504/IJART.2011.041486.

Szell, M., Thurner, S., 2010. Measuring social dynamics in a massive multiplayer online game. Soc. Netw. 32, 313–329. http://dx.doi.org/10.1016/j.socnet.2010.06.001.

Taylor, T.L., 2012. Raising the Stakes: E-Sports and the Professionalization of Computer Gaming. MIT Press.

Thrift, N., 2005. Knowing capitalism. Knowing Cap. 1–65. http://dx.doi.org/10.4135/9781446211458.

Trepte, S., Reinecke, L., Juechems, K., 2012. The social side of gaming: how playing online computer games creates online and offline social support. Comput. Hum. Behav. 28, 832–839. http://dx.doi.org/10.1016/j.chb.2011.12.003.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications, vol. 1. Cambridge Univ. Press, pp. 116. http://dx.doi.org/10.1525/ae.1997.24.1.219.

Watts, D.J., 2007. A twenty-first century science. Nature 445, 489. http://dx.doi.org/10.1038/445489a.

Yang, P., Harrison, B., Roberts, D.L., 2014. Identifying patterns in combat that are predictive of Success in MOBA games. Proc. Found. Digit. Games 2014. pp. 1–8.

Zhong, Z.J., 2011. The effects of collective MMORPG (massively multiplayer online role-playing games) play on gamers' online and offline social capital. Comput. Hum. Behav. 27, 2352–2363. http://dx.doi.org/10.1016/j.chb.2011.07.014.