Dynamic Mechanisms forming Networks of MOBA Matches: Understanding Overlaps, Topology, Resilience and Diffusion in a Gaming Community.

Pedro C. Vieira
Department of Computer Science, Faculty of Sciences
University of Porto
Rua Campo Alegre 1021/1055, 4169-007
Porto, Portugal
pedrocvieira@fc.up.pt

João Afecto
Department of Computer Science, Faculty of Sciences
University of Porto
Rua Campo Alegre 1021/1055, 4169-007
Porto, Portugal
joao.afecto@fc.up.pt

ABSTRACT

This work provides insight into the construction of networks of MOBA games and produces an analysis of the dynamics and topology of such a network by studying a sample of a network of League of Legends players. By understanding the dynamics and topology of the network it is possible to tune mechanisms involved in the process that generates the network in order to enhance the resilience and immunity of the network. These properties account for a better player experience by providing a smoother experience e.g. by avoiding the propagation of scam attempts. Furthermore, the insight into the dynamics and topology of the network allows us to better understand the mechanisms involved in the generation of the network e.g. the matchmaking system, and to have a better view of the behaviour patterns of players e.g. how many players have a massive match presence.

1. INTRODUCTION

Online gaming is a trend that has been popular for many years and, with the development of powerful hardware with increasing availability to the public, it is expected that it so remains for many years to come. A popular type of game is Multiplayer Online Battle Arena (MOBA). Examples of games in such genre are Smite, Dota 2, Pokemon Unite and the object of study of this work, League of Legends.

The multiplayer online battle arena (MOBA) represents a subgenre within the realm of strategy video games, wherein two teams of players engage in competitive gameplay on a predetermined battlefield. Each participant assumes control over a single character possessing a unique array of abilities that are enhanced throughout the game, thereby influencing the strategic direction of their team [15].

As stated previously, our work will revolve around League of Legends (LoL). The developer of LoL, Riot Games offers ten distinct server regions [2] for players based on their location. We will work only with data from the region Europe West (EUW1). The data used was extracted using the official Riot-API [1].

The main goal of our study is to understand how the players connect. We want to understand what is the topology of the network of players. Some factors that might influence the network are the matchmaking system, together with the different habits and frequency of matches per player.

2. DATA

2.1 Strategy

As mentioned in section 1, the data used was obtained using the official Riot API. In order to manage rate limits imposed by the API, an automated query-and-store system was built. From the search conducted, the distribution of ranks peaks either at Silver or Gold [4; 25]. Hence, it was decided to start the search at a rank that had the most people. Hence, the search was started at a random player from the Gold rank. The total player count for EUW1 is around 34 million (with inactive players) [25]. However, due to limitations on the time needed to acquire data, we only were able to obtain data for 313673 players, originating from 50000 matches. The strategy of acquiring data is displayed in list 2.1.

1. Start from a random player at the specified rank (Gold).
2. Add it to the queue.
3. Store that player information.
4. For all players x in the queue:
   (a) Get their most recent ten matches.
   (b) For each player y in the ten matches store them if they are not stored.
   (c) If the player y is not in the database add him to the queue.
   (d) Remove player x from the queue.

From list 2.1, we can see that a natural graph arises from the matches between players. The graph will be undirected and connects players that played with each other in at least one of the most recent ten games of one of the players. Since players can have their most recent ten matches played at a different time window, there will be enough variance to secure diversity in the network.

Strategy 2.1 uses breadth-first search (BFS) as the search mechanism to track players as they appear in matches. In

the tree, T, produced by the algorithm, it exists both levels of matches and players. The structure of the levels is given by interpolated levels of players and matches, starting at a player level and finishing at a match level. Figure 1 shows an example of a tree resulting of the application of the algorithm 2.1. The squared nodes represent a player, and the circular nodes a match. Since we have two different types of nodes in T, it will exist two different branching factors for T. The branching factor for the number of games originating from a player is given by m and represents the number of matches we expect to see given a player. The branching factor for the number of players a match originates, p, describes the number of players we expect to see in each match. In figure 1, $m = 3,\ p = 5$. Note that in all levels except for the root, only $m - 1$ new matches are produced and $p - 1$ new players appear. Hence the effective branching factors to all minus the root level are $m_e = m - 1$ and $p_e = p - 1$. Based on the aforementioned finding, it becomes possible to deduce the contact rate, represented by the variable d, for an arbitrary player within the context of the study. We can overlook the root since it accounts only for a small deviation from a regular tree. We expect a player to contact with $d = m_e \times p_e$ new players.

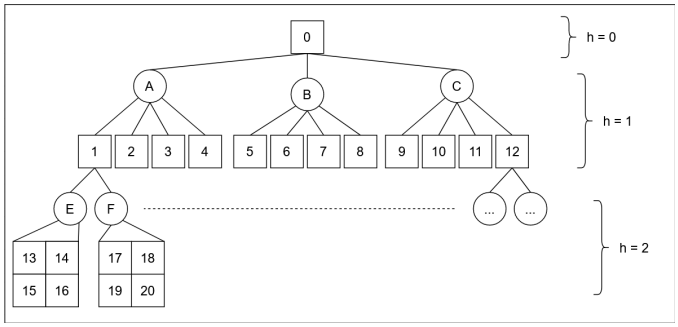

Figure 1: Representation of the tree constructed by strategy 2.1. Circular nodes represent a match and square nodes a player. Each level of the tree except for the root is composed of a layer of each node type.

After the aforementioned framework, our objective is to ascertain the extent of player overlap within the dataset comprising the 50000 extracted matches. Specifically, we aim to validate the hypothesis that each match does not exclusively introduce new players. Leveraging the coefficients introduced in the preceding paragraph, we can compute the expected number of players under the assumption of a complete absence of overlap among them.

Firstly we need to define sufficient relations to compute said overlap based on the regularity of the tree. We first define such relations assuming that the branching factors are all equal, in this case, equal to the effective branching factor. The number of players at a given height/level $h \geq 0$ is $(m_e \times p_e)^h$. At any level $h \geq 1$, we have $m_e\,(m_e \times p_e)^{h-1}$ matches. Having the foundation presented in the last paragraph, we can evolve the relations to achieve greater accuracy in the calculations by adding the branch that is left unattended when we base ourselves only on the effective branching factor. Restricting the height h to start at one instead of zero, we can define a general equation (equation (1a)) to give the total number of players at a level h.

$$P(h) = (m_e\, p_e)^h + p_e^h\, m_e^{h-1} \qquad h \geq 1 \tag{1a}$$

$$P_{\text{total}} = 1 + \sum_{h=1}^{\infty} P(h)$$

In light of starting from an initial height of one, in equation (1a), we must include an additional player, thereby ensuring the accuracy of the final count.

For the total of matches at a height $h \geq 1$, the formula is given in equation (2).

$$M(h) = m_e\,(m_e\, p_e)^{h-1} + (p_e\, m_e)^{h-1} \qquad h \geq 1$$

$$M_{\text{total}} = \sum_{h=1}^{\infty} M(h) \tag{2}$$

Since at $h = 0$, they are no matches, there is no need to define an extension of the equation for matches to account for referred height.

Knowing that we pulled 50000 matches, solving the equation for the length of the summation, or maximum height of the tree, we can derivate the expected number of players if no overlap exists between said players. Equation (3) shows the equations used. To advance from equation (3b) to equation (3c), it used a closed-form formula for the partial sum of a geometric series. Since it is possible to see the most recent ten matches and each match as ten players, we have $m_e = 9,\ p_e = 9$.

$$M_{\text{total}} = \sum_{h=1}^{t} m_e\,(m_e\, p_e)^{h-1} + (p_e\, m_e)^{h-1} \tag{3a}$$

$$\equiv M_{\text{total}} = \sum_{h=0}^{t-1} m_e\,(m_e\, p_e)^{h} + (p_e\, m_e)^{h} \tag{3b}$$

$$10\left(\frac{1-81^h}{1-81}\right) = 50000 \tag{3c}$$

$$\equiv h \approx 2.9353$$

$$P_{\text{total}} = 1 + \sum_{h=1}^{3} (m_e\, p_e)^h + p_e^h\, m_e^{h-1} \tag{3d}$$

$$P_{\text{total}} = 597871$$

The approximation maximum height would be 2.9353. This result comes from a real solution of the equation. However, we are dealing with an integer number for the variable h. The closest integer approximation of the solution concerning the summation after transforming back to $h = 1 \rightarrow t$ is 3. Hence, in equation (3d), we assume a maximum height equal to three.

From equation (3d) we reach a value of 597871 players obtained when following strategy 2.1. In light of this result, we see that it exists an overlap of players of $1 - 313673/597871 \approx 47.44\%$. The fact that we approximated the height to an integer solution means the number of games is over-represented. Knowing that 66430 games exist when the height of the search tree is equal to three we can produce a more accurate statistic. Calculating the excess games, 16430, and admitting independence between games we reach the conclusion that $16430 \times 10 = 164300$ players participated in the excess

games. Taking this excess into account, the new overlap statistic is $1 - 313673/(597871 - 164300) \approx 27.65\%$. This result means that at least, 27.65% of players are repeat and at most 47.66% of players are repeated.

2.2 Processing

Upon successful data storage, the JSON files were subjected to parsing utilising the OpenSearch search engine, leading to the subsequent construction of the network. The network resulting from this construction had some undesired properties. It had a large number of cliques resulting from the mechanics of a match. That is, every player involved in a match connects to any other player in the match. This fact permits cliques of ten people to appear more than expected. A symptom of this topology is that such a network had an average clustering coefficient of 0.889. With such a high clustering coefficient and following a probabilistic approach, overall, in the network, the probability of two random nodes connecting is 88.9%. In order to address this concern, we opted to condense the network by reducing the identified cliques within it. The initial stage entails the computation of all cliques present in the network, employing the implementation provided by `networkit` based on the algorithm proposed by David Eppestein and Darren Strash [21]. The second stage was to create a new network with nodes representing each clique found in the original network. Every node in the new network is a super-node in the sense that it represents multiple nodes. For every node i of the new network, one edge is added from i to j if any of the nodes of the original network represented by i connect to any nodes of the original network represented by j. The condensed network built does not contain multi-edges or self-loops by construction. The obtained network has 44587 nodes and 779615 edges. Algorithm 1 gives a brief description of how the condensation is achieved.

Algorithm 1 Condensation of the network based on cliques

```
Require: cliqueMap = MapEachCliqueToId()
Require: nodesInClique = MapNodesToCliqueId()
Require: cliqueList = ListAllCliques()
 1: G ← EmptyGraph()
 2: nodes ← ∅
 3: for clique in cliqueList do
 4:     for srcMNode in cliqueList do
 5:         nodes ← graph.ListEdges(srcMNode)
 6:         srcSNode ← getClique(srcMNode)
 7:         for pDstMNode in nodes do
 8:             if pDstMNode not in clique then
 9:                 cliquesDst ← nodesInClique[pDstMNode]
10:                 for dstSNode in cliquesDst do
11:                     if dstSNode ≠ srcSNode then
12:                         G.AddEdge(srcSNode, dstSNode)
13:                     end if
14:                 end for
15:             end if
16:         end for
17:     end for
18: end for
```

Repeating lines 7,8 and 9 of algorithm 1, changing `nodes` for `cliquesDst` and making other small adjustments makes so that the `srcMNode` has the opportunity of potentially have more connections [1]. These steps called *condensation reach* govern the number of connections a node can have. Lines 10 to 12 only concretize the connections made by the condensation reach.

As the algorithm is presented, each super node formed by a clique can "see" and connect to direct connections of any of the nodes it direct represents. These zero-order connections allow a super node to reach a maximum of connections equal to $m_e \times p$. Furthermore, since the algorithm has one iteration of the stage "condensation reach", it can see and connect to first-order nodes. One example is displayed in figure 2. In the said figure, the algorithm is looking at what cliques it should consider connecting to the node representing the clique in white. It is currently considering the node marked with a 1. If no steps of condensation reach were applied, the algorithm would only consider the cliques limited by the brown nodes (clique A and B) as connections to the white clique. If "condensation reach" was applied once (configuration used), cliques up to the purple nodes (irst-order nodes) would be considered. Hence, applying one step means that, besides A and B, cliques C and D would be considered. Applying two steps would yield the additional cliques E and F up for consideration. The formula for the number of connections available for a super node is given by equation (4).

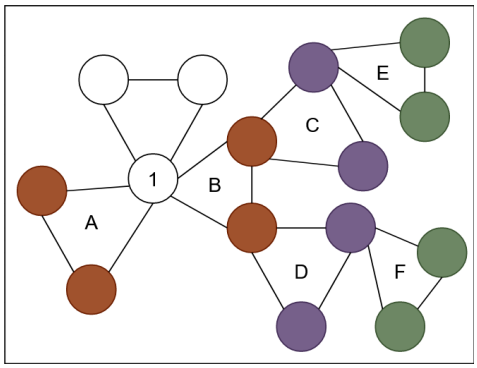

Figure 2: Representation of zero-order (brown), first-order (purple) and second-order (green) connections possible using varying number of condensation reach stages.

$$\text{Max Connections} = \kappa(t) = (m_e \times p) \times (m_e \times p_e)^t \quad (4)$$

Furthermore, a bound can be given on the number of connections. Using the interval of possible overlap presented earlier, we expect that the maximum number of possible connections a super node can have in the configuration used is no less than 4155.

Figure 12 showcases a visual representation of the network, with the layout acquired through the ForceAtlas2 algorithm. Noteworthy observations from the visualization include the emergence of distinct structures positioned towards the lower section, as well as certain components attempting to disengage from the central region on the right-hand side. Furthermore, a preliminary formation of two central "clusters" of nodes is apparent, delineated by an "island" configuration isolating them from the surrounding central nodes. It

[1] More details in the code attached with the report.

is important to note that the current layout, generated by ForceAtlas2, as specified, has yet to achieve convergence. Consequently, it is plausible that certain structures may represent transient artefacts arising during the formation of more definitive network structures.

2.3 Correction To Maximum Connections

The theory established in the past sections holds except for a correction factor δ. This factor is immeasurable. Hence, we cannot approximate it analytically. Consequently, it cannot be accurately estimated *à priori*. Even though the connections are always bidirectional, it is possible for players to "forget" they played with another player. Let us give an example. Let us have a player x and a player y. Let player x and player y play together. Now let us allow player y to not play enough so that the match between him and x remains in its history. As for player x, let him be active enough so that its match history does not show a connection to y. When querying player y, a connection will be added to x. Since the network is undirected by nature and design, x will also connect to y. However, x will continue having the number of connections predicted in earlier sections. The quantity $\delta(x)$ measures the number of players that node x has that it "forgot", such is the case of y, and are not accounted in the previous calculations. Hence, any player may have more connections than the ones theorized before.

This correction reflects a change in $\kappa(t)$ in the following way:

$$\kappa\prime(t) = \kappa(t) + \sum_{u \in U} \delta(u) \tag{5}$$

where U represents the set of nodes collected during $\kappa(t)$. Hence, the results presented above constitute an absolute lower bound on the number of connections.

The existence of the δ factor makes the time complexity of the algorithm hard to quantify. A over-pessimistic approach that does not take into account the number of condensation reach steps is $O(c \cdot n)$ where n is the number of nodes in the graph and c the number of cliques of the graph.

3. TESTS AND METHODOLOGY

From the network resultant from the application of the procedure explained in subsection 2.2, we will compute a series of statistics in order to evaluate the main characteristics of the referred network.

The main goal of this work is to understand the dynamics and consequently, the topology of the network in order to better understand the real-world system behind it.

We will start with the inclusion of a null model and tests to refute such a null model. We will then explore possible mechanisms that could possibly generate the network. We will start by testing the Barabási-Albert model and explore some extensions to encompass evolving scenarios.

Finally, we will conduct a small practical example of network robustness by inverse percolation theory in the context of network science. In the context of network science we mean we want to study the behaviours of a network system when some of the nodes or links are not available [28]. We will also produce small practical experiments relating to diffusion.

The software used was Networkit [35] and graph-tool [33] for general statistics calculation, HoloViz [34] for visualization, and the RAPIDS suite [36] for multiple tasks parallelizable through GPU.

Table 1: Basic statistics for NoM.

Number of Nodes	Number of Edges
44587	779615
Avg. Clustering Coefficient	Avg. Path Length
0.338	3.657
Pseudo Diameter	Avg. Degree
10	34.97 ± 35.5

All calculations were computed using an AMD EPYC 7443 Processor (16 cores, virtualized) with 128 GB of main memory and a NVIDIA GeForce RTX 3090 (24GB VRAM) GPU.

4. NETWORK MODELS

4.1 Null Model

Within this section, our focus directs towards investigating properties associated with the topology of the network. The initial step involves the establishment of a null model. In this case, we select an Erdos-Renyi ($G(N, p)$) random network as the chosen null model. Our objective is to refute the null hypothesis and the identification of an alternative model that can be deemed the primary mechanism responsible for generating the Network of Matches (NoM).

The main basic statistics about NoM are presented in table 1. Following descriptions and results presented in [10], we expect that for the null model, the average degree to be $\langle k \rangle = p(N-1)$, the diameter $d_{\max} \propto log\, N / log \langle k \rangle$ and the average clustering coefficient for a node i and for the whole network to be $C_i = \langle k \rangle / N$. The distribution of such a network is binomial. However, when the subject is sparse networks, $\langle k \rangle \lll N$, the degree distribution can be approximated with a Poisson distribution of the form $p_k = exp(-\langle k \rangle)\langle k \rangle / k!$.

Concerning the estimation of the network diameter, it is notable that, in the majority of networks, it provides a more accurate approximation of the average path length ($\langle d \rangle$) between any two nodes within the network. This better approximation arises because $d_{\max}$ is frequently influenced by the presence of unusually long paths. In contrast, the average path length ($\langle d \rangle$) mitigates such fluctuations by encompassing the average of all paths, thereby attenuating the impact of possible outliers.

The generation of an Erdos-Renyi (ER) random network necessitates the specification of two key parameters: the number of nodes denoted as N and the probability of linkage represented by p. The determination of the number of nodes, N, is derived directly from the network under investigation. The probability of linkage, p, can be inferred from the relationship to the average degree of the network. Hence, we have $N = 44587$ and $p = \langle k \rangle / (N-1) \approx 0.0007$.

If NoM was random, having in mind that the $\langle k \rangle \approx 35$, we would have an average clustering coefficient equal to $p = 0.0007$. Knowing that the clustering coefficient for a particular node is also equal to 0.0007, it would be very unlikely for two nodes that connect to a third one to connect between themselves. The average clustering coefficient for NoM is 0.33. This result means that the expected probabil-

ity of two neighbours of a randomly selected node connecting is orders of magnitude larger than in a random network.

One observation noteworthy of mention is that since NoM is a condensed network from the original network with a high number of cliques, the clustering coefficient does not represent a statistic regarding a unique player but rather a statistic about a clique. Given three matches m_1 , m_2 and m_3. Each match has ten players. Match m_1 and m_2 connect to m_3 through some player. A possible interpretation of the clustering coefficient in the referred setting is the probability of m_1 and m_2 connecting through any player in each of the matches. In other words, it represents the probability of any player from a match playing with any player from one other match, knowing that they have played a common person before. A visual comparison of the distribution of the clustering coefficient for NoM and for a network generated by ER is available in figure 3. In said figure we can see the regularity of the clustering coefficient of a random network, while observing the more broad nature of NoM. We also see that NoM has overall a higher clustering coefficient per node. As for the random network, the majority falls between 0.01 and 0.001, with emphasis on the 0.001, as we can see by the more demarked blue spheres. This result is a respectable approach to 0.0007.

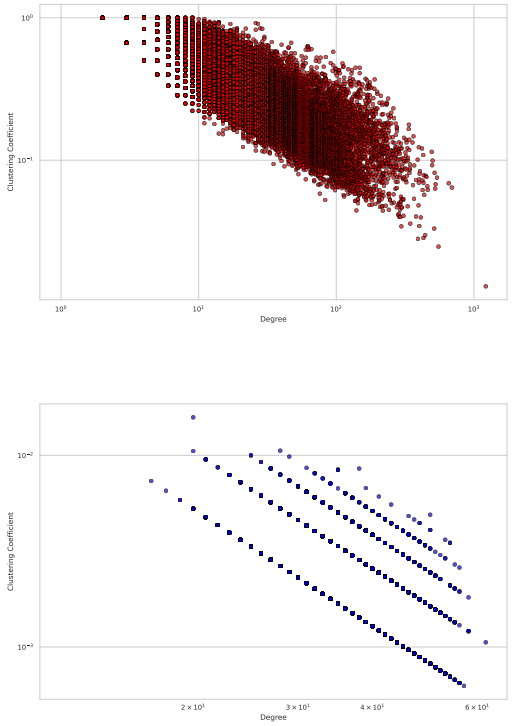

Figure 3: (TOP) Clustering coefficient for distribution for all nodes of NoM. (BOTTOM) Clustering coefficient for distribution for all nodes of an ER Network made to match NoM. Data is presented in log-log scale.

Another measure of a clustering coefficient is the global clustering coefficient or transitivity. Computed as the ratio of three times the number of triangles over the total amount of triplets. Even though it is not equivalent to the average clustering coefficient, it is another way of measuring the triadic closure probability described above.

Considering an Erdos-Renyi random network, the corresponding set of vertices is defined as $V = \{1, 2, \ldots, n\}$. To compute the expected count of triangles in this random network it is necessary to treat the number of triangles as a random variable, denoted by X. Let X represent the observed count of triangles within the graph G. The maximum number of subsets of size three G can have is $\binom{n}{3}$. For a subset of size three to be a triangle it has to have all the links between the nodes. The probability of such a thing occurring is p^3. Knowing we may observe $\binom{n}{3}$ each with probability p^3, and knowing that all triangles are equally likely to appear, the $E[X] = \binom{n}{3}p^3$.

Expanding on the perspective presented above, in order to calculate the expected value for the number of three-node chains, we have to consider the two possibilities that can originate a three-node chain. The first one is inside a triangle. The second one is as a "pure" chain. We already calculated the expected number of triangles $E[X] = \binom{n}{3}p^3$. For each triangle, we have three chains. Hence, for a random variable representing the observed chains inside a triangle, the expected value of one such variable is $E[Y] = 3\binom{n}{3}p^3$. Furthermore, we can have a chain with just three nodes and two links. We have $\binom{n}{3}$ groups of three nodes. We want to get at least two links. Hence we would have $\binom{n}{3}p^2(1-p)$. In subsets of three nodes we have $\binom{3}{2}$ combinations of links, totalling $\binom{n}{3}p^2(1-p)\binom{3}{2}$. For random variable Z representing the observed number of chains in subsets of size 3, we have $E[Z] = \binom{n}{3}p^2(1-p)\binom{3}{2}$. The total expectation would be $E[Y+Z]$. Using the linearity of expectation we have:

$$E[Y+Z] = E[Y] + E[Z] = 3\binom{n}{3}p^3 + \binom{n}{3}p^2(1-p)\binom{3}{2}$$
$$= \binom{n}{3}\left(3p^3 + p^2(1-p)\binom{3}{2}\right)$$

Hence, for an ER random network, the transitivity is given by equation (6).

$$3 \times \frac{E[X]}{E[Y+Z]} = 3 \times \frac{\binom{n}{3}p^3}{\binom{n}{3}\left(3p^3 + p^2(1-p)\binom{3}{2}\right)} \tag{6}$$

For a random network with the characteristics of NoM, the expected transitivity score would be 0.0007 as given by equation (6). In fact, from the equations above, it is possible to conclude that these two coefficients will always be the same for an ER random network. The transitivity coefficient for NoM is characterised by 3046462 triangles and 54588999 triplets, giving a score of 0.167.

The degree distribution of a random network can, as explained, be approximated by a Poisson distribution with high accuracy in the case of sparse networks. The network in question is sparse, manifesting $< 0.1\%$ of the total amount of edges possible. Hence, if random, the degree distribution would follow a Poisson distribution characterised by $\lambda = \langle k \rangle$. The expected degree distribution for a random network following the Poisson distribution, the degree distribution obtained by an ER network, and the distribution for NoM are displayed in figure 4. All axis are set to a logarithmic scale.

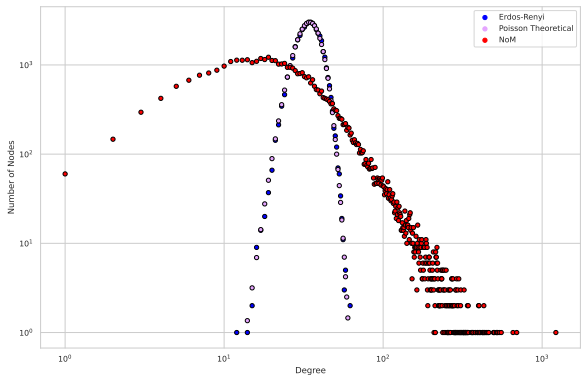

Figure 4: Degree distribution plotted on log-log scale for NoM, a network generated by the ER model with parameters to match NoM and the theoretical approximation of the distribution of such network.

Upon examining figure 4, it becomes evident that the degree distribution of the Network of Matches (NoM) distinctly deviates from the other two distributions. This significant dissimilarity, coupled with the previously established properties, provides compelling evidence to reject the proposed null model. Hence, we conclude that the sample of the network produced by the matchmaking system is not random.
Notwithstanding the substantial existing evidence regarding the inadequacy of the null model, a final comparison will be conducted focusing on the normalized variants of the closeness, betweenness, and eigenvector centralities. Figure 5 illustrates the division of the three centralities based on the network type. The standard methodology was employed to normalize the closeness and betweenness centralities. Regarding the eigenvector centrality, the Euclidean norm was utilized to establish a valid centrality measure, which attains its maximum value at $\sqrt{1/2}$ [3].
Observing the betweenness centrality measures depicted in figure 5, it is evident that the values attributed to NoM span a considerable range, varying from relatively small magnitudes to orders of magnitude larger. Additionally, a notable pattern emerges, wherein a substantial majority of nodes exhibit negligible centrality values, while only a light fraction of nodes possess significantly higher values. Consequently, the distribution of this centrality measure exhibits a pronounced heavy-tailed characteristic. In comparison, the distribution of values for the ER network is much closer to a symmetric distribution with only a light tail at the higher values. Furthermore, the range of values originating from the centrality of the random network is much tighter, whereby we notice the lack of large values as in the case of NoM. The presence of nodes with a large betweenness centrality in NoM is most likely due to the presence of nodes with a very high degree, as can be seen in figure 4. In the same figure, we notice the absence of such nodes in the random network.
The closeness centrality displayed in figure 5 shows a very close resemblance between the distribution of both networks. The distributions are similar both in shape and in range of values, only with NoM perhaps presenting excess kurtosis

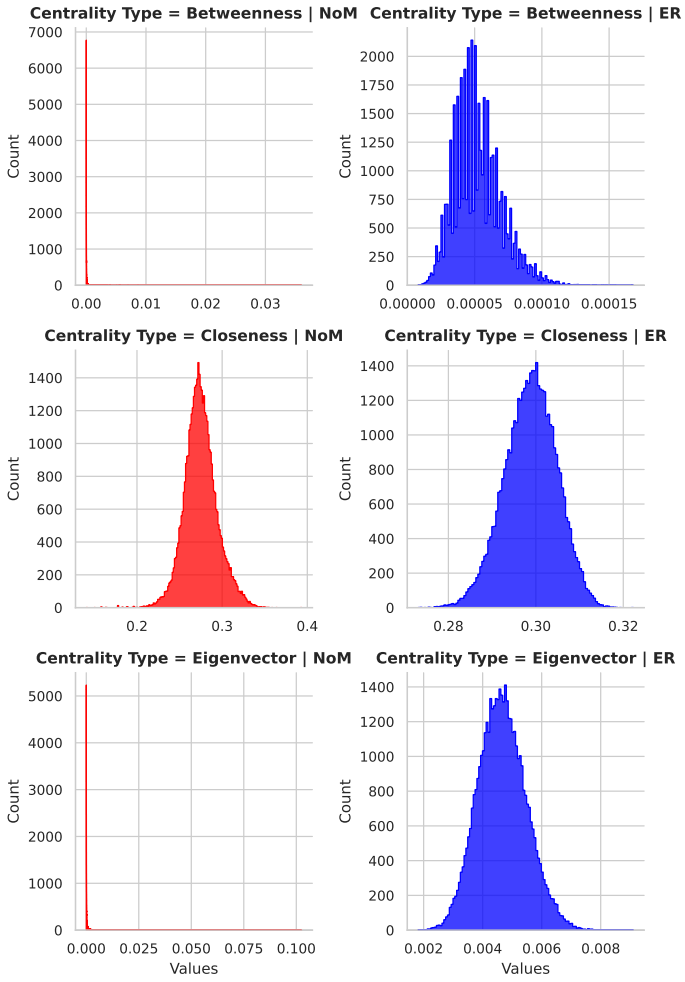

Figure 5: Normalized centralities for NoM and ER network with parameters matching NoM.

concerning the random network. This result may indicate that the overall distance of a node to all others is not that different from one network to another.
The eigenvalue centrality presents a behaviour in all similar to the betweenness centrality. The fat tail presented in the distribution for NoM most likely belongs to the nodes with a very high degree and to the ones that, despite having few connections, such connections are directed to important nodes.
The distinctions noticed from the centralities are most likely due to the topology of NoM, particularly the existence of highly connected nodes. As for the similarity in the closeness centrality, it gives us an indication of the closeness of the nodes in the network, stating that the difference of distance from one node to all others is not very dissimilar on both networks. A deeper understanding of this matter can be done by looking at the average path length of the networks. Table 1 shows that the average path length for NoM is 3.657. The expected approximation for a random network is of the magnitude of $\frac{log\, N}{log\langle k\rangle}$. Hence, for the ER network with parameters with respect to NoM, the average path length would be 3.012. This result indicates that the two networks have

an average path length of the same magnitude. Such a result coupled with the one regarding the closeness centrality leads us to believe that the nodes with very high degrees are not enough to shorten the distance between nodes. Effectively leading to metrics close to the random model.

4.2 Beyond the Null Model

Section 4.1 allowed us to refute the hypothesis that the NoM originates from a univariate purely random process. Furthermore, it offered clues about where we should go in order to further refine our hypothesis about the process of generating the network in question. The plot of the degree distribution presented in figure 4 raises suspicion of a power law being the distribution that generates the degree distribution of NoM. Given these indications, an experimental approach was conducted to explore the suitability of fitting specific distributions known to manifest in real network systems. Such distributions yield degree distributions akin to the observed pattern [10].

We followed the discrete formalism to guide the fitting procedure since a node has a degree that is always a positive integer. Consequently, the primary contender (power law) for serving as the underlying distribution responsible for generating the degree distribution of NoM is characterized by the following probability mass function [10; 16]:

$$p(x) = \frac{k^{-\alpha}}{\zeta(\alpha, k_{\min})}$$
$$\zeta(\alpha, k_{\min}) = \sum_{n=0}^{\infty} (n + k_{\min})^{-\alpha}$$

where ζ is the generalized zeta function and $k_{\min}$ represents the minimum degree at which the data is drawn from the power law distribution. Futhremore, since said distribution diverges at 0, we must have $k_{\min} > 0$. Since NoM has only one connected component, there will be no need to cut any node from the estimation.

The fitting procedure followed by maximum likelihood estimation and the goodness-of-fit tests were based on the Kolmogorov-Smirnov (KS) statistics [16; 6].

Besides the power law distribution, we fitted four more distributions, all following the procedure described in the paragraph above. An exponential distribution, since typically the definition of heavy-tail distribution is that it is not exponentially bounded [7]. It was also fitted a log-normal distribution, stretched exponential and a truncated power law distribution, since these are common in real network systems [10; 16].

Table 2 shows the values for the parameters encountered for each distribution. Exceot for the exponetial distribution, all distributions passed the initial KS test. In order to understand what would be the best distribution among the possible ones, a normalized loglikelihood ratio test was conducted according to specifications of Clauset et al. [16]. The log-normal distribution presented a statistically better fit for the data when compared to the pure power law, but presented a worse result than the truncated power law and stretched exponential. As for the comparison between the stretched exponential and truncated power law, the former performed worse than the latter. All distributions exhibited a superior goodness of fit in comparison with the exponential distribution Notably, despite achieving statistical significance (p-value at 5% confidence level) for all the comparative analyses conducted, the observed ratio remained relatively modest in magnitude except for the exponential distribution. Furthermore, even though small, the truncated power law, stretched exponential and log-normal distributions have a higher degree of freedom than the power law, thus a fitting advantage. Hence, we may look into these results carefully. Furthermore, it is well documented the difficulty in distinguishing between these distributions in practice [6; 16; 10; 26]. It is also advised that in cases of statistical doubt, remember the basic principles of the domain-specific generative mechanisms. With those, one can form a basis to decide which distribution is better fitted.

Table 2: Fitted parameters for the degree distribution of NoM. The $k_{\min}$ for NoM that optimizes fitting is 48.

Distribution	Power law	Log-Normal	T. Power law	Exponential	Stretched Exponential
α	3.12	——	2.47	——	——
λ	——	——	0.005	0.0267	11.27
μ	——	1.71	——	——	——
σ	——	1.19	——	——	——
β	——	——	——	——	0.30

The log-normal and truncated power law distributions can be classified as an interpolation of an unbounded distribution, such as the power law, and a bounded one, such as the exponential distribution. Even though for large values these interpolated distributions have a tail that decays exponentially or faster, depending on their parameters, they can be used to fit an unbounded distribution.

The truncated power law can be defined similarly to the power law. The main difference is that it incorporates an exponential coefficient as can be seen in equation (7).

$$p(x) = \frac{\lambda^{1-\alpha} k^{-\alpha} e^{-\lambda k}}{\zeta(1-\alpha, \lambda k_{\min})} \tag{7a}$$

According to Barabási [10], the truncated power law exhibits the behaviour of a power law up to a high-degree cutoff point where the exponential decay of the tail takes over. For $k \ll 1/\lambda$ the distribution follows the power law with degree α, and once the high-degree cutoff point is reached at around $k \geq \lambda$, the distribution exhibits an exponential cutoff.

As for the log-normal distribution, it arises when multiple independent random variables are multiplied together. This factor means that the distribution is normal in $log\, k$. Even though this distribution is common in finance it is also used frequently to fit network degree distributions [16; 10]. Since this distribution is inherently continuous, in order to estimate it for the discrete formulation, an approximation by rounding was used [6].

The stretched exponential is similar to the truncated power law distribution but has a fractional power law in the exponential.

Figure 6 displays the approximated probability density functions (PDFs) of the fitted distributions for the power law, log-normal, and truncated power law. However, the PDF for the exponential distribution is omitted due to its notably inferior fit. Notably, in the case of the stretched exponential, the fitted PDF closely resembled that of the truncated power law to such an extent that distinguishing between them became intricate.

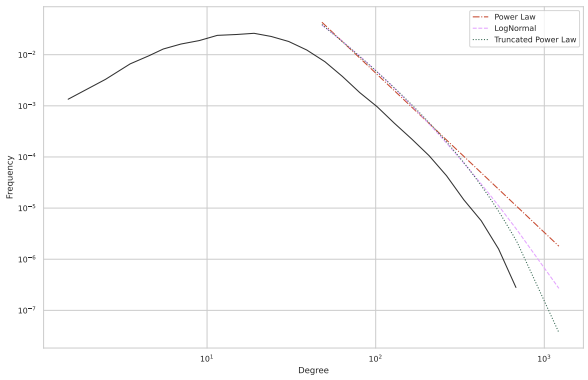

Figure 6: Comparison of different distributions fitted to the degree distribution of NoM (black). Used $k_{\min} = 48$.

From the analysis of figure 6, we can see a similar behaviour between the fitted log-normal and truncated power law. Both exhibit a high-degree cutoff point at which the number of nodes with a large degree decays exponentially or nearly exponentially. Both these distributions seem to equally be a good fit for the part of the distribution that follows a power law. Hence, the slight advantage of the truncated power law might come from the fact that it seems to approximate better the decay of the tail of the distribution. Nevertheless, the "pure" power law distribution appears to exhibit a slightly improved fit within the region of the distribution where a power law is presumed to hold, compared with the others. Hence, the primary drawback of the power law distribution appears to stem from its inability to accurately approximate the behaviour beyond the point of high-degree cutoff.

The shape of the degree distribution of NoM is commonly observed in real systems [10]. The low-degree saturation point marks the region where the network has fewer nodes with low degrees than the expected power law, typically marked by $k_{\min}$. Furthermore, the high-degree cutoff zone marks a zone from which the number of nodes with a high degree appear in less quantity than what we would expect in a power law. The high-degree cutoff also limits the size of the node with the highest degree. One possible explanation for the low-degree saturation and the high-degree cutoff is the fact that the network represents a sample of the complete network. Hence, it is natural that some, perhaps inexistent, variation in the network exists that is not present in the complete system. Furthermore, the low-degree saturation may be also influenced by the overall high activity of players, making low degree nodes less likely to appear than the estimated by the BA model. As for the high-degree cutoff, the limitation of ten matches per player may be a very important factor in such a behaviour of the network.

A network characterized by a degree distribution that adheres to power law is commonly referred to as a scale-free network [10]. It is worth noting that neither the presence of a low-degree saturation nor a high-degree cutoff serves as an impediment to the scale-free nature of the network. Scale-free networks demonstrate robustness in the face of low-degree saturation [10; 19], and there exists even a generalization framework to accommodate such cases [19]. As for the high-degree cutoff, it limits the divergence of the second moment of the degree distribution. However, the presence of such a cutoff is an indicator that additional possibilities exist after the cutoff region, possibly indicating the need to further study such a period. It is not an indicator of deviations from a scale-free system.

The aforementioned network type exhibits distinctive phases based on the exponent α. When the power-law exponent α assumes a value of 2, the degree of the largest hub displays linear growth in proportion to the system size. This phenomenon results in the formation of a "hub and spoke" configuration within the network. Hence, the average path length does not depend on the size of the network. When the power law exponent is between 2 and 3, the first moment $\langle k \rangle$ is finite, but the second moment diverges, resulting in networks with a large range for the degree of a node. This fact allows an "ultra small world" where the distance scales proportional to $log\,(log\,N)$. If $\alpha = 3$, the average path length is proportional to $log\,N/log\,(log\,N)$. When the exponent is larger than 3 the second moment no longer diverges. Despite the existence of highly connected nodes, they are not sufficient to impact significantly the distance between nodes. Hence, the average path length comes back to scaling proportionally to $log\,N$ as was the case with a random network.

In the case of the network in question, the second moment of the degree distribution is large compared with the first moment. However, the average path length scales like a random network as was explained in section 4.1. The estimated exponent for α in the case of the "pure" power law is 3.12. If we consider the truncated power law, the exponent estimated for the power law period is 2.47. These results lead us to make the hypothesis that the system may be scale-free and fall on the ultra-small world scale or be scale-free and fall in the zone where the second moment converges. These two hypotheses are not contradictory. In fact, the best hypothesis comes from joining them. The existence of a high-degree cutoff may be part of the reason why we get a coefficient above three. That is because we lack the expected number of hubs, and the hubs with the expected high degree, the average path length becomes proportional to a system in such a zone. However, when we take into account the cutoff as not being part of a power law, the exponent drops to 2.47 (truncated power law) hence, we empirically observe such a high second moment. Hence, depending on what parts of the degree distribution we consider when fitting, we might get a valid, but different result. These different results also push for different interpretations. Hence, we theorize that either option is valid for NoM. That is, one may interpret it as following a truncated power law, and the other may interpret it as following a pure power law with an exponent greater than 3. If interpreted as a "pure" power law we might blame the experiment for the anomalies at the start and end of the system. In this case, for the complete system, from the analysis of the sample in question, we believe it would follow a power law with an exponent in the vicinity of 3, perhaps slightly larger, perhaps slightly smaller. If we incorporate said anomalies into the system we might adopt the truncated power law and search for explanations for the anomalies as well as mechanisms to replicate them. Reasons based on properties of the real system for the high-degree cut off and low-degree saturation, were already theorized earlier before.

One may question the characteristics of NoM based on the regularity of the network constructed by unique players. That is, due to the limitations given by the API, one player is limited to having a degree in the range of 9 to 90. However, this problem is solved by condensation of the cliques within a network, as explained in section 2.

Nonetheless, the limitation imposed on the number of nodes limits the growth phenomena ("the rich get richer") [6]. Hence, from a point onward, it is known that the power law will not grow as expected due to reaching the maximum stress of the system (nodes cannot gain more edges). Hence, a truncated power law may be a better fit in those cases. However, due to the size of the network, and the bound established in section 2, we believe that the maximum stress point of the system is not at play here.

4.3 Possible Generation Mechanisms

The first and most well-known mechanism used to generate scale-free networks was presented by Barabási and Albert [8] and is based on the presence of growth and preferential attachment. The proposed generative process has two mechanisms (1) the addition of a new node at each timestamp t that will connect to m nodes in the system - growth; (2) each node that arrives at each timestamp t will connect to other nodes already present with probability proportional to the degree of each node - preferential attachment. One common constraint is to have $m \leq m_0$, where m_0 is the initial number of nodes in the system. However, this formulation can be relaxed using $m_0 = 0$ and allowing for new nodes to connect to as many nodes as possible until the threshold m. This model is known as Barabási-Albert (BA) model and has the characteristic that it produces a network with $\alpha = 3$ regardless of m, m_0 and t [10]. In the previous section, a hypothesis was put forward regarding the degree distribution of the Network of Matches (NoM). Specifically, it was postulated that NoM may follows a power-law distribution characterized by a degree exponent slightly greater than 3. Adjusting the parameters following equations (8), approximating the NoM should be possible.

$$N = t + m_0 \tag{8a}$$

$$E = m_0 + m \times t \tag{8b}$$

In equation (8), $m_0 = 0$, N equals the number of nodes in the network and E is the number of edges. We want to discover the number of links used in the growth step of the model. Solving the system yields $m \approx 18$. Figure 7 shows the approximated degree distribution for the NoM and a network generated with $m = 18, m_0 = 0, N = 44587$. The black vertical line represents $k_{\min}$.

Analysing figure 7 shows that the degree distribution does fit the degree distribution of NoM. Further analysis shows that the tail of the generated network is "heavier" than the tail of NoM. This fact is possible to connect to the previously mentioned high-degree cutoff experienced by NoM. It is important to note that while observing a satisfactory fit of a model with a degree of 3, such fit does not provide conclusive proof that the power law governing NoM possesses a degree of 3. It does present an additional challenge to rejecting the hypothesis. Only in the event of the fitted network yielding significantly unfavourable outcomes would it be possible to draw further conclusions. However, one indication that we can conclude from the simulation is that the BA model can-

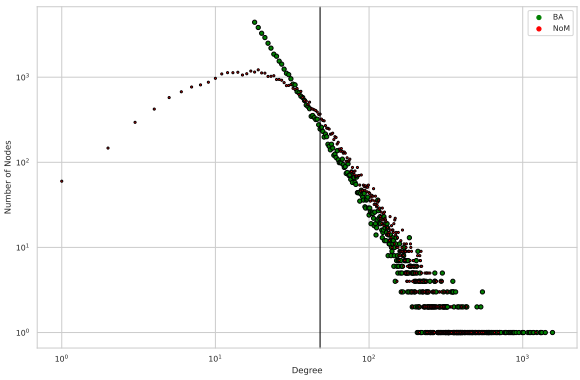

Figure 7: Degree distributions for NoM and a network generated by BA.

not cause either low-degree saturation or high-degree cutoff. One potential explanation for the presence of a high-degree cutoff, as previously discussed, is the inherent limitations of the experimental setup. Another plausible explanation lies in real-world phenomena. The phenomena causing high-degree cutoff can be modelled through non-linear preferential attachment, specifically a sublinear variant [10]. However, accurately testing for such a property proves to be exceedingly challenging within a static scenario. Regarding the low-degree saturation, like the high-degree cut off, it can derive from the limitations imposed by the experiment. However, it can also be modelled by extensions to the BA model.

The Barabasi-Albert (BA) model exhibits several additional limitations worth noting. Firstly, the model assumes that all nodes grow at an equal rate, resulting in older nodes consistently maintaining an advantage in attaining higher degrees. Consequently, new nodes cannot surpass older ones in terms of degree. Secondly, the BA model does not incorporate the concept of node or link disappearance, which can significantly impact the dynamics and evolution of real-world networks. One other limitation is the restriction of the exponent to 3. Nonetheless, this mode presents a steeping stone for the seach of better suited model.

4.4 Dynamic Models

The generation of a power law distribution with an exponent distinct from 3 can be achieved by employing alternative generative models. The Bianconi-Barabási (BB) model serves as an illustrative example in this context [12]. This model can be viewed as an extension of the Barabasi-Albert (BA) model, incorporating growth and preferential attachment mechanisms alongside a fitness parameter η that affects how a node acquires new neighbours. The impact of fitness on preferential attachment can be quantified using equation (9).

$$p(i) = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{9}$$

The distribution, $\rho(\eta)$ generating the fitness for each node can in theory be any real function. However, not all func-

tions will give a useful result. If ρ is a constant function the BB model generates a network with the same characteristics as the BA model [12]. One case of particular interest for the current experiment is when ρ is a uniform function. In this case, the network generated will have $\alpha \approx 2.255$. In the section regarding the fitting of a distribution to the degree distribution of NoM, the truncated power law estimated $\alpha = 2.47$. Therefore, considering the theoretical framework of a truncated power law, it is plausible to suggest that the generation of a network resembling the Network of Matches (NoM) can be achieved through the adoption of the Bianconi-Barabási (BB) model. In order to simulate such a generative model, it would be ideal to understand the range and the values the fitness function may assume. Typically in a real system, the fitness reflects the perception of the importance of each node in comparison with the others. One possible way to determine the fitness function would be to construct it uniformly based on the ranks of the games (e.g. average ranks of all participating players or the average of all players of all matches a node represents). However, at time of writing, we do not possess that information. Furthermore, evaluating the fitness values of nodes in a static network is challenging beyond the scope of this work. The only works known to us regarding studying a network for traces of a fitness function in the BB model consist of either simulating a dynamic system from the static sample or capturing a network mid-formation [27]. Barabási [10] presents a good description of this method.

Many other extensions of the BA model exist. One such example is the concept of initial attractiveness. The initial attractiveness A affects each node equally, hence being an added constant in the probability distribution of preferential attachment. This extension enhances the probability of new nodes receiving connections. Hence, we will have fewer nodes with a low degree as opposed to what should be expected according to a power law. This result means that the network will experience a small-degree saturation. As for the influence of A in nodes with a high degree, the small and constant nature of A make so that the impact in nodes with a high degree is diminished, not affecting the system in this region [19]. The saturation point is given by $k < A$. It is also expected for the degree exponent to increase above 3 [20; 10].

Another extension of the BA model is the introduction of links within the pre-existing nodes [9]. The probability distribution of linkage between the pre-existing nodes is given by equation (10).

$$p(i, j) = (A_i + B_i\, i)(A_j + B_j\, j) \tag{10}$$

By putting $A_i = A_j = 0$ the mechanism that guides new links between pre-existing nodes is preferential attachment. In this case, α takes ranges between 2 and 3. Putting $B_i = B_j = 0$ the internal attachment is random. This creates a higher homogeneity between the degrees of the nodes. Hence, the degree will be > 3.

The extensions discussed for the Barabasi-Albert (BA) model primarily focus on augmenting or modifying the mechanisms of edge and node addition. However, it is important to acknowledge that in real-world systems, the presence of nodes and links may also involve their disappearance. One straightforward approach to incorporate this element involves expanding the node addition step to include the possibility of removing a node with a probability of r [20; 10]. If $r < 1$, then the system continues growing, now more homogenous due to node and link removal. If $r = 1$, the network remains with a constant size. Following the absence of growth, the network will lose its scale-free property. For $r > 1$, the network will decline and disappear. More complex mechanisms in the node removal can lead to more complicated dependencies such as a phase transition between scale-free and exponential systems.

Several other extensions have been proposed for the Barabasi-Albert (BA) model, including accelerated growth [20; 22; 24; 10] and ageing [18; 10]. However, based on our assessment, it would be unexpected for these extensions to account for the distinctive characteristics observed in the Network of Matches (NoM).

The choice of a particular extension for network modelling depends on the specific characteristics of the network under investigation. Each extension may be more suitable for certain types of systems, while others may be better suited for different contexts. To comprehend the underlying dynamics of a network, it is crucial to grasp the dynamics that drive its formation. Understanding network topology is an added result of understanding its dynamics.

In previous sections, we theorized that both the low-degree saturation and high-degree cutoff might come from limitations in the experiment. If this was the case, the network, if completed, would follow a pure power law. This explanation is plausible. At the end of section 4.2, we noted that one other possibility for the high-degree cutoff was the existence of non-linear preferential attachment. This is a possible explanation but disregards the existence of a low-degree saturation. In this section, we explored extensions to the original BA model. The Bianconi-Barabási (BB) model allows the use of fitness, an attractive parameter to model a ranked matchmaking network. Furthermore, a uniform distribution for the fitness function allows for a power law with an exponent close to the one estimated by the truncated power law. Hence, one possibility is to allow both the low-degree saturation and high-degree cutoff to be responsible for the limitations of the system and use a BB model to estimate the power law within those deviations. As emphasized earlier, comprehending the dynamics of a network is a crucial prerequisite for gaining insights into its topology. With the range of extensions discussed earlier, we have options that enable the direct incorporation of both low-degree saturation and high-degree cutoff into the model. The key challenge lies in identifying the most appropriate extension that can effectively capture the characteristics of the real-world system. Achieving this entails bridging the gap between the model and reality, a task that necessitates significant domain expertise and understanding.

Initial attractiveness seems a very plausible characteristic of the real-world system. Like in the fitness function theorized in the BB model, the rank system causes the matchmaking system to not be completely biased to high-degree nodes. Hence, it does not obey preferential attachment entirely. Furthermore, as matches get older they will be removed from the player history, hence, it is natural that node and link removal exist in the complete system.

In light of the considerations made in the last paragraph regarding the behaviour of the real-world system in question, if we must attribute the anomalies registered at the start and end of the distribution, it makes sense to use initial at-

tractiveness and node and link removal. Previous research [11] shows that these dynamics can produce a degree distribution with low-degree saturation and high-degree cutoff. In theory, such distributions are better fitted with a stretched exponential. Hence, perhaps why the stretched exponential resulted in a fit almost as good as the truncated power law.

5. EXPERIMENTS

In this section, we undertake a series of small-scale experiments aimed at assessing the robustness of the network through the lens of inverse bond and site percolation. Additionally, we explore various diffusion models that draw parallels to the dynamics observed in the real-world system. These experiments serve to probe the network's resilience and evaluate the efficacy of different diffusion mechanisms in capturing the intricate dynamics at play.

5.1 Inverse bound and site percolation

Bond percolation and site percolation are processes studied in physics [28]. Simplified, in the context of network science, site percolation relates to the addition of nodes adjacent to others in a network with probability p. As for bond percolation, it relates to the addition of edges between already present nodes with probability p [28]. The inverse percolation system consists of doing the opposite. Typically, it conveys removing either node (site) or edges (bonds) from the network. Both bond and site percolation provide insights into the connectedness and robustness of networks by studying the emergence of connected clusters or the disintegration of the network.

We will start with testing the robustness of the network to the removal of nodes. The results derived from the studies conducted by Molloy and Reed [30] and further elaborated upon by Barabási [10] reveal a special aspect of the robustness exhibited by scale-free networks. Specifically, it has been demonstrated that for networks with a power-law degree distribution characterized by an exponent $\alpha \geq 3$, the critical threshold of robustness is determined solely by the minimum degree $k_{\min}$, independent of the network size. This finding underscores the inherent resilience of scale-free networks and highlights the significance of the minimum degree as a key determinant in shaping their robustness properties. One of the theories developed in section 4.3 states that the power law followed by the complete real-life system may have a degree in the vicinity of 3. Furthermore, the power law with the best fit has a degree of 3.11. Hence, it might make sense to test the robustness of the network to significantly different $k_{\min}$. If the observed low-degree saturation is indeed rooted in system dynamics rather than being an artefact of experimental limitations as postulated in section 4.4, a possibility for enhancing the robustness of the network can be design based on $k_{\min}$. Specifically, by considering the potential influence of the initial attractiveness, which is likely a primary mechanism underlying this saturation phenomenon. Hence, it becomes feasible to fine-tune the degree distribution to achieve heightened robustness characteristics.

A network of matches connected through shared players should provide valuable resilience in the face of potential attacks. This robustness can be particularly beneficial in scenarios where an attacker seeks to disrupt the player experience by disturbing the historical records and replays of previous gameplay or by undermining the anti-cheat traceback mechanisms. By leveraging the interconnectedness of the network, it becomes more challenging for malicious actors to isolate a match or compromise the integrity of the system as a whole.

Figure 8 shows the effect of inverse site percolation for NoM and other BA networks with comparable degree exponent but different $k_{\min}$. The x-axis shows the number of nodes in the network and the y-axis the size of the largest connected component. The order of removal is by nodes with higher number of vertices first, simulating an attack starting on the most highly connected individuals. Furthermore, the attack starting with the most connected individuals is compared against a null model. The null model consists of randomly removing nodes in the network.

One noteworthy observation is that NoM suffers from a high-degree cutoff. However, the overall results should would, even though some fluctuations may exist when compared to a "pure" scale-free network such as the networks used for comparison.

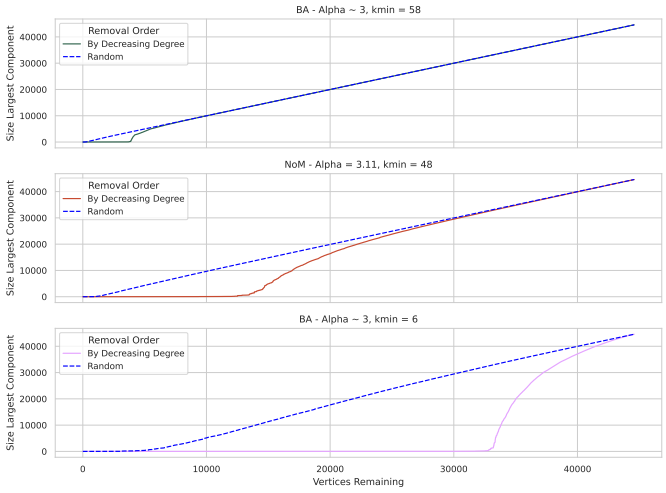

Figure 8: Evolution of the size of the largest component as nodes are removed of the network. A null model consisted of random removal and a attack model consisting of removing nodes starting from the ones with a higher degree are compared.

Looking at figure 8, we can see that the null model poses no threat to the overall connectedness of any of the networks. For this type of attack, the critical value c = nodes removed / total nodes, in order for the network to suffer a severe loss of connectedness is close to 1. This ratio means that, for a random attack, we expect that, for each node removed, the size of the connected component to decline by one. For $k_{\min} = 58$ and an attack starting at the most connected individuals, we see a decline faster than the null model after 10000 nodes remaining. A critical point is fast approached by having the largest component have a size of practically zero when around 5000 nodes remain. When we reduce $k_{\min}$ to 48, the disparity between the null model and the attack based on the highest individual increases. Before having 10000 nodes remaining, the size of the largest component is nearly zero (technically the size is one, and only zero when no more nodes exist). When $k_{\min} = 6$, the critical value hap-

pens much sooner. Prior to reaching a node count of 30000, the size of the largest component approaches zero. Moreover, as the minimum degree ($k_{\min}$) decreases, the critical threshold not only occurs earlier but also induces a more pronounced decline.

One academic utility of these tests is to further cement the theory of network following a heavy-tail distribution and not a exponential one. The disparity in vulnerability between scale-free and exponential networks with respect to the type of attack has been documented in the literature [5; 14]. Scale-free networks are typically more resilient to random failures than exponential networks, while being more vulnerable to attacks by highest degree first. Whether this observation is perceived as increased vulnerability or enhanced security is subjective and open to interpretation.

We now proceed with a view at inverse bond percolation. Similar results for predicting a critical point for the removal of edges exist [10; 17; 14]. Nonetheless, we move with the experiment. Figure 9 shows the results of the conducted experiment. Unlike in the site percolation experiment, where the number of vertices is the same by construction, in this experiment, the number of edges is different in order to keep the number of vertices the same. All results hold regardless. However, a direct comparison is not possible.

Edges with a higher weight with will be removed first when modelling an attack. In order to evaluate the weight of an edge, it was used the product of the degree of the nodes that it connects.

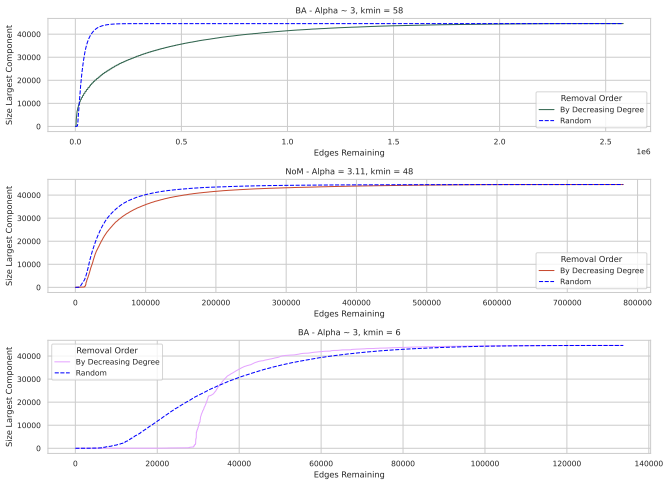

Figure 9: Evolution of the size of the largest component as edges are removed of the network. A null model consisted of random removal and a attack model consisting of removing edges starting from the ones with a higher weight are compared. Weight was defined by the multiplication of the degree of the nodes that an edges connects.

Looking at figure 9, we can see that for $k_{\min} = 58$, the decline occurs much earlier for an attack on the most important edges. Nevertheless, in contrast to the null model, the decrease in the component size of the network under investigation is noticeably more gradual. The null model exhibits a sustained maximum component size until a critical threshold is reached, after which it experiences a sharp decline. As the minimum degree $k_{\min}$ decreases, the onset of the decline in the size of the largest component is delayed. However, once a critical value is reached, the descent becomes more pronounced and exhibits a steeper decline. The opposite happens for the null model.

5.2 Diffusion Models

In this section, we will conduct small practical experiments regarding diffusion models. The field of diffusion of information on complex models has theoretical results that illuminate and guide practical experiments. Only in very complex systems, more often regarding dynamic transition probabilities, that are outside of the analytical frameworks described in the classical literature, the simulations gain enough room to be relevant. Theoretical results on the used models and others can be viewed in [29; 31; 13; 32].

In this experiment, we will be looking at some simple practical situations where diffusion models can be used to predict the spread of information. We will use two models, the SIR and the SIRS model. Let $X_i = \{\alpha_0, \alpha_1, \alpha_2\}$, where α_x denotes a different state that a node may encounter itself on. If a node i is at a specific time t in state α_x than $X_i(t) = \alpha_x$. In the SIR and SIRS compartmental models, α_0 represents a susceptible state, α_1 the infected state and α_2 the recovered state. The probability transitions are given as follows:

$$X_i(t) = \begin{cases} p_i(\alpha_1) = 1 - \Pi_j (1 - \beta_{i,j})^{A_{i,j}\delta(X_j(t),\alpha_1)} & X_i(t) = \alpha_0 \\ p_i(\alpha_2) = 1 - \mu_i & X_i(t) = \alpha_1 \\ p_i(\alpha_0) = 1 - \gamma_i & X_i(t) = \alpha_2 \end{cases} \tag{11}$$

The first branch of equation (11) means that if a node is in state α_0, it will pass to α_1 with probability $p_i(\alpha_1)$. Following the same principles, it transits to α_2 with probability $1 - \mu_i$ and to α_0 with probability $1 - \gamma_i$ when in α_1 and α_2 respectively. Equation (11) reflects the SIRS model. When evaluating the SIR model, the last branch does not exist. Hence, once a node transits to α_2, it stays there forever.

The experiment conducted concern attempts of scams that can occur during a match. These attempts of scams can be either to perform identity theft, buying scams or even money fraud. Typically scams diffuse between scammed individuals and their connections. It is common to use friendship connections, but it is also possible to use contact connections, for example, matches in common. For the diffusion probabilities, we will use the results from a survey from the European Commission published in 2020 [23]. According to the refereed survey, the probability of being in contact with a scam is 56%. However, as far as we are concerned, the report is ambiguous regarding which percentage falls for the scam. Hence, we will make an educated guess, based on experience, and say that an online player would fall for a scam ten times less than the rate at which contact with scams happen. This assumption makes the probability of falling for a scam 5.6%.

In the SIR model, the state α_0 regards the nodes that have a normal probability of falling for a scam (5.6%). State α_1 regards nodes that fall for the scam. Nodes in this state will propagate the scam to their connections. Nodes in state α_1 can undergo a learning process that promotes increased caution, leading to a transition to state α_2, where they become resistant to scam activities. The probability of this transition happening is 38%. It is based on the percentage of people that alter their habits after being exposed to scams

[23]. The results are displayed in figure 10.

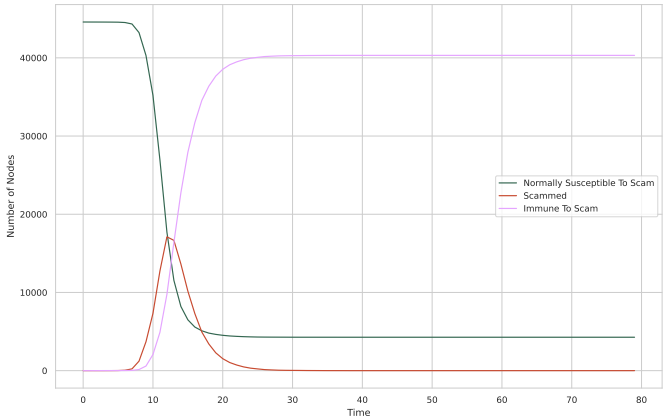

Figure 10: Evolution of the state of the nodes of NoM according to the diffusion model SIR. The green line corresponds to state α_0, the brown one to α_1 and the pink one to α_2.

Looking at figure 10, we can see that the number of people that never suffered a scam attempt decreases drastically and reaches the lowest point at around 20-time steps after patient zero. We also observe that the peak of scammed individuals occurs around time-step number 10. However, due to the high recovery rate of the nodes, the diffusion of fraudulent activities reaches zero slightly after time-step 20. Consequently, there will be nodes within the network that will never encounter any scam attempts.

The second diffusion model admits that nodes do not become immune after becoming more aware of scams but enter a hardened period. During this period, nodes are hyper-aware and do not fall for any fraudulent activity. However, they will eventually become lenient again, resulting in returning to the initial state. The probability of transition from the hyper-aware state to the initial state was deemed equal to the probability of becoming hyper-aware. The result of the simulation is presented in figure 11.

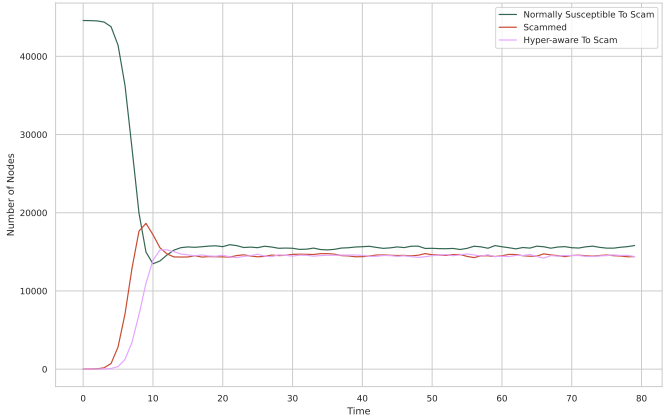

Figure 11: Evolution of the state of the nodes of NoM according to the diffusion model SIRS. The green line corresponds to state α_0, the brown one to α_1 and the pink one to α_2.

Analysing figure 11, we can see a similar drop in nodes in state α_0 as in figure 10. A similar peak in scammed individuals also exists. However, since nodes can transit from the hyper-aware state to the initial state, we no longer observe the transition of the state of the system into an immune state (fixed-point). Contrarily, we see that scammed individuals keep appearing. We see a stabilization of the number of scammed nodes and hyper-aware nodes. This finding implies that while the scam epidemic remains unresolved and persists within the system, the number of individuals being scammed at each time step is effectively controlled. One noteworthy observation is that contrary to the SIR model, there is no guarantee that one node will never experience a scam attempt. In figure 11, we see a stabilization of the nodes in state α_0, but there are no guarantees that the nodes in state α_0 are always the same.

6. CONCLUSION

This work starts with the construction of a network of players of League of Legends. The nodes corresponded to players and the edges between nodes connect players that had a match together. The study was initiated by outlining the data acquisition process and network construction. Through the conducted analysis, it was determined that the player overlap within the network ranges from 27.65% to 47.66%. This finding indicates that the addition of each new match is anticipated to introduce approximately six to eighth new players to the network.

After a short analysis of the network described in the last paragraph, it was understood that the characteristics of the referred network were "too regular to produce insight into the system". Hence, an algorithm for the condensation of the network based on existing cliques was devised. This algorithm allows the representation of cliques of nodes through a super node. Furthermore, it allows for a flexible view of the original network by allowing connections of arbitrary depth to be taken into account when constructing edges in the new network. The new network constructed was nicknamed *Network of Matches* (NoM).

After the elaboration of NoM, an analysis was conducted in order to understand the dynamics and topology of NoM. This analysis starts with the use of several metrics in order to refute a null model stating that the NoM originates from a random Binomial process. After the refutation of the null hypothesis, a series of tests regarding dynamic models was conducted. These tests will produce further insights into the dynamics and topology of NoM. The main results of the referred tests indicate that NoM is scale-free. Furthermore, it was concluded that NoM either follows a power law with a coefficient around 3 and the anomalies registered at the start and end of the degree distribution are due to the limitations on capturing the complete system, or it follows a stretched-exponential or truncated power law and mechanisms such as node and edge deletion, and initial affinities are at play. A less likely option is for the network to follow a Bianconi-Barabási process.

The final experiments conducted aimed at improving the resilience of the network by tunning $k_{\min}$ and studying the propagation of fraudulent activities in the network. The tests conducted regarding resilience were based on inverse percolation theory (bond and site), and the propagation of fraudulent activities on diffusion models. The results ob-

tained here reinforce the theory developed in other works.

APPENDIX

A. EXTRA IMAGES

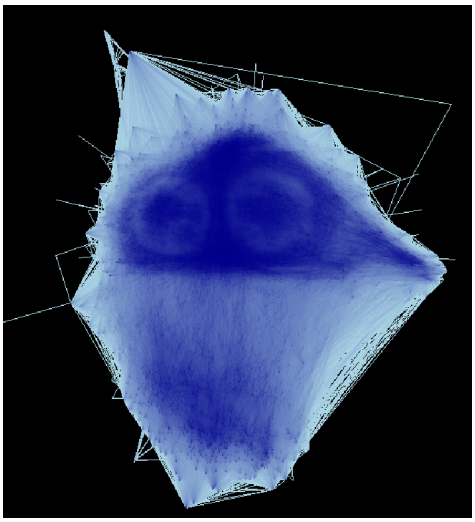

Figure 12: Visual representation of NoM. Dark blue points represent nodes and light blue edges.

B. REFERENCES

[1] Developer page riot games. Accessed on 2023-06-13, https://developer.riotgames.com.

[2] Support forum league of legends. Accessed on 2023-06-13, https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201751684-League-of-Legends-Regional-Servers.

[3] Eigenvector-centrality — a node-centrality? *Social Networks*, 22(4):357–365, oct 2000.

[4] S. Ahn. League of legends rank distribution: Where do you rank? Accessed on 2023-06-13, https://riftfeed.gg/more/league-of-legends-rank-distribution-where-do-you-rank.

[5] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[6] J. Alstott, E. Bullmore, and D. Plenz. Powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777, may 2014.

[7] S. Asmussen, S. Asmussen, and S. Asmussen. *Applied probability and queues*, volume 2. Springer, 2003.

[8] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science (New York, N.Y.)*, 286(5439):509–12, oct 1999.

[9] A. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614, Aug. 2002.

[10] A.-L. Barabási and M. Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[11] H. Bauke, C. Moore, J. B. Rouquier, and D. Sherrington. Topological phase transition in a network model with preferential attachment and node removal. *The European Physical Journal B*, 83(4):519–524, Oct. 2011.

[12] G. Bianconi and A.-L. Barabási. Competition and multiscaling m evolving networks. In *The Structure and Dynamics of Networks*, volume 54, pages 436–442. Princeton University Press, 4 edition, mar 2001.

[13] M. Boguá, R. Pastor-Satorras, and A. Vespignani. Epidemic spreading in complex networks with degree correlations. In *Statistical Mechanics of Complex Networks*, pages 127–147. Springer Berlin Heidelberg, 2003.

[14] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471, Dec. 2000.

[15] A. Cannizzo and E. Ramírez. Towards procedural map and character generation for the MOBA game genre. *Ingeniería y Ciencia*, 11(22):95–119, July 2015.

[16] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. 2007.

[17] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628, Nov. 2000.

[18] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks with aging of sites. *Physical Review E*, 62(2):1842–1845, aug 2000.

[19] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of Growing Networks with Preferential Linking. *Physical Review Letters*, 85(21):4633–4636, nov 2000.

[20] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of Growing Networks with Preferential Linking. *Physical Review Letters*, 85(21):4633–4636, nov 2000.

[21] D. Eppstein and D. Strash. Listing all maximal cliques in large sparse real-world graphs, 2011.

[22] M. J. Gagen and J. S. Mattick. Accelerating, hyperaccelerating, and decelerating networks. *Physical Review E*, 72(1), July 2005.

[23] Ipsos. Scams survey on and fraud experienced by consumers, Jan 2020. Accessed on 2023-06-21, https://commission.europa.eu/system/files/2020-01/factsheet_fraud_survey.final_.pdf.

[24] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, Oct. 2000.

[25] R. Kamberovic. League of legends player count: Here are the stats. Accessed on 2023-06-13, https://riftfeed.gg/more/player-count.

[26] A. Klaus, S. Yu, and D. Plenz. Statistical Analyses Support Power Law Distributions Found in Neuronal Avalanches. *PLoS ONE*, 6(5):e19779, may 2011.

[27] J. S. Kong, N. Sarshar, and V. P. Roychowdhury. Experience versus talent shapes the structure of the web. *Proceedings of the National Academy of Sciences*, 105(37):13724–13729, Sept. 2008.

[28] M. Li, R.-R. Liu, L. Lü, M.-B. Hu, S. Xu, and Y.-C. Zhang. Percolation on complex networks: Theory and application. *Physics Reports*, 907:1–68, apr 2021.

[29] A. L. Lloyd and R. M. May. How viruses spread among computers and people. *Science*, 292(5520):1316–1317, May 2001.

[30] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, Mar. 1995.

[31] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B*, 26(4):521–529, Apr. 2002.

[32] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, Aug. 2015.

[33] T. P. Peixoto. The graph-tool python library. *figshare*, 2014.

[34] P. Rudiger, J. Signell, J. A. Bednar, , Andrew, J.-L. Stevens, C. B, J. Samuels, , Todd, T. PEDOT, S. V. D. Oord, J. Mease, I. Virshup, G. Corona, D. Hermes, H. Curtis, and A. Graser. holoviz/hvplot: Version 0.5.2, 2020.

[35] C. L. Staudt, A. Sazonovs, and H. Meyerhenke. Networkit: A tool suite for large-scale complex network analysis, 2014.

[36] R. D. Team. *RAPIDS: Libraries for End to End GPU Data Science*, 2023. https://rapids.ai.