Methodology of Data Science: League of Legends as a Case Study

Alyssa M Adams[1,2,3,*]

[1]Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe
[2]Algorithmic Nature Group, LABORES, Paris, France
[3]Department of Physics, Arizona State University, Tempe
[*]amadam15@asu.edu

ABSTRACT

Although living systems compose a large part of the world around us, they still remain elusive to a complete and consistent mathematical model, despite such models being previously successful in the physical sciences. The underlying mechanisms of life remain poorly understood across all scales of biological organization. On the other hand, technology is providing us with an overabundance of data about digital social systems. Since this data is a representation of one level (or a few levels) of biology, can it be used to understand living and complex systems more generally? A large movement within complex systems science aims to achieve just that. Using League of Legends data as an example, this paper explores the limitations of technological data and how it might be used to understand living systems more broadly. While it is true that much of the data provided by technological systems is helpful for understanding the dynamics of that particular system, the data itself is already missing several key details that are crucial to how living systems operate. A new framework for thinking about systems, including living ones, is introduced as a baseline for future discussion of collecting better data. Although not a rigorous theory, it is a mental bridge to begin traversing the space between current physical laws as we know them and the mechanisms that drive living systems.

Disclaimer: All data collection, processing, and visualization have been previously published in a paper called *Understanding Real-world Open-ended Systems: A League of Legends Adventure*, which was accepted for publication in the International Journal of Design & Nature and Ecodynamics in April 2017. This paper serves at a critique of that paper and therefore contains many of the same figures.

Introduction

Arguably one of the most challenging scientific endeavors at the forefront of 21st century research is the quest to understand living systems and how they differentiate themselves from non-living systems. We currently have few insights into how living systems might *quantifiably* differ from their non-living counterparts, as in a mathematical foundation to explain away our observations of evolution, emergence, innovation, and organization. Development of a theory of living systems, if at all possible, demands mathematical understanding of how data is generated, collected, and changes over time, not unlike current well-established scientific disciplines. After all, living systems are comprised of physical elements: atoms, charges, and masses. It seems that it is only a matter of time before someone comes along and fits a mathematical model that bridges the gap between our observations of living systems and current scientific models. Given a collection of observations on some level of scale— chemical, quantum, astronomical— a mathematical model is invented to explain and, more importantly, predict those observations. Through this process, most scientific disciplines are formed.

According to many philosophers who agree with Alfred N. Whitehead, the folly of these various (largely) disconnected scientific disciplines is the omission of perfectly valid observations and measurements in order to satisfy a rigid mathematical approach. Human thoughts such as "What do yellow tulips mean?" and "What was I thinking when I ordered that nasty taco" have no place in current mathematical models. Although Fourier transformations can be easily applied to classical music in order to analyze which frequencies are most common, it lends no understanding to how it makes a person *feel* or what thoughts it elicits to the listener. The elimination of such details (details that are common and entirely vital in living systems) is hindering our understanding of what life is and what not-life is. As Whitehead is so insistent about, the mere observation of facts as positivists is a nonsensical approach to understanding Nature. Why should scientists only be concerned with plain descriptions of Nature without probing their underlying meanings? Common criticisms to this claim include using teleology, or giving 'will' or meaning to objects. Here, it is important to draw a boundary where 'meaning' is being applied: Here, studying meaning does not imply that all objects have a 'meaning', such as the 'meaning of an electron moving through a magnetic field.'

Instead, the idea of 'meaning' is given to living entities, much like we humans draw meaning from printed symbols. An RNA molecule has a well-defined function in the appropriate context of a cell, which is not to exclude 'meaning of the RNA to the cell'. **Subjectivity, rather than pure objectivity, of the laws of nature is the main thesis of this paper.**

Current mathematical models that describe the precession of physical entities thorough space-time, or any combination of such entities, are not sufficient to understand the phenomenon that are more common to experiencing everyday life as humans[1–3]. And yet, we remain largely ignorant to the importance of experience and sensing. I want to stress that although these terms are generally used to describe the human experience, it should not exclude the most general notions of experience and sense as possible. For example, how does a cell 'sense' RNA? Because of this failure, and perhaps others, we currently lack a coherent theory general enough to encompass important aspects of life, which are the crux of commonplace thoughts and feelings among humans that drive our societal processes.

Technology

On the other hand, technology in today's (respectively more) globalized society is providing us with an overabundance of data[4]. More than ever, human activity is being tracked with finer and finer resolution in terms such as video and text. At any given time, we are able to count the number of global Tweets produced by people who "follow" NBA players. We are also able to track the number of changes made to George W. Bush's Wikipedia page over time[5]. Because of these trends, we have seen recent work on understanding large-scale biological processes shift focus to studying innovative technological and social systems, such as the evolution of patents[6,7] and online social systems[5,8,9]. However, given the pitfalls of the methodological approaches mentioned earlier, is approaching this data with the same scientific mindset a fruitful path of inquiry to understanding living systems? Technological data represents human behavior, which arguably is centered entirely on sense and experience. Without taking these two crucial factors into consideration, processing modern technological data may be of no assistance in understanding living systems if processed in the traditional sense.

The purpose of this paper is to explore various avenues to approach this problem. Here, a full set of data is presented, along with a desired interpretation of the data, and is followed by a lengthy discussion of how to move from data to interpretation. This exercise is critical for considering living systems on other levels of organization; cellular, multicellular, chemical, etc. Since the larger goal is to understand living systems across all levels of its organization, we carefully presume that this analysis can be used as a surrogate for all other levels under appropriate abstractions. We therefore search for ways to make generalizations in a manner that is much different than has been done in traditional scientific disciplines in order to preserve details that are crucial to our understanding of life.

Example Data: League of Legends

The dataset presented is a representation of human interactions in a virtual world; an online competitive video game called League of Legends. Suppose our question is this: Given current data that has been collected from the game, is it possible to predict (in some form or another) what *future data* will look like? Often, many businesses are interested in these types of questions to guide their business strategies[10]. Is it possible to tease out a model that explains how the game's data is generated? Can we use that model to predict future data and be prepared for possible future trends?

League of Legends (League) is an online community of players where millions of people play with and against each other on a virtual battlefield, pitting 140+ in-game characters against one-another in team battles. The game is developed by Riot Games (Riot) and is regularly modified in code every week or two. The interactions between players within the game and regular external interventions by Riot lead to a complex dynamical system. One characteristic of this complexity is that successful player strategies never stay static; A player's strategy that wins many matches today may become a terrible strategy next week. The game's complexity contributes to the game displaying many 'life-like' properties that resemble features seen in other complex biological systems, as a result of human-to-human interactions. In particular, the game as a whole (collection of all people who play League, the people at Riot, and software/hardware interfaces) updates itself over time according to its own history and its current state, making it a self-referential system.

In every instance of a single match in League, ten players are randomly partitioned into two teams of five: the "Blue" and "Red" team. The two teams battle against each other on a virtual battlefield, as shown in 1. Each player picks a single in-game champion to play throughout the entire match and the player uses that champion to compete against the enemy team for resources on the map. Two instances of the same champion are not allowed in a single match. Resources, used to level up champions throughout a match, are collected around the map and can be earned by killing enemy champions or by interacting with the map in various ways.

The goal of the game is for one team to destroy the other team's base. This requires a high level of cooperation and trust within individual teams of players. The more matches a single player wins over time, the higher the player climbs up in skill. League has a skill ranking ladder that sorts players into different tiers by skill. These tiers are used to randomly select players into teams for matches, such that all players in a single match are of equivalent skill level.

Figure 1. Summoner's Rift, the virtual battlefield ("map") where both teams fight and collect resources. Team bases are on opposite corners of the map. Spatial areas of the map are labeled in white according to their colloquial names. For a sense of scale, one champion "Teemo" is shown on the map. Players who want to play "Teemo" often do so in the "Top Lane."

As Riot updates League's code, most changes are aimed at making the game more fun for the players[11]; listening to player needs and wants is the differentiating goal of Riot as a game company. The majority of Riot Games' changes are based on feedback from the players, therefore the game's code evolves according to the opinions of the players, which are generated by interacting with League.

First Abstractions

There is no question that the game is an open-ended system since its development is driven by human innovation and creativity, which is already assumed to be open-ended[12,13]. For these reasons, League makes a good example case of modern living systems described in some way by a set of data. Because it is comprised of living entities (humans) and elements constructed by living entities (computers, the Internet, hardware, etc.), it is one of many levels of biological organization. Understanding general properties about this level of organization could be useful for understanding other levels of organization. At best, properly assessing the abundance of technological data could provide a highly detailed assessment of one level of biological organization (if used properly), which could be generalized to understand other levels of biological organization.

To attain these generalizations, some level of abstraction is necessary, with a loss of information deemed not important. One such abstraction is presented here although many other representations of the game are equally valid, given the type of abstraction is highlighting aspects of the game to be studied. The game can most simply be partitioned into two interacting subsystems: The players and the game developers. Both these subsystems update each other's behavior much like a biological entity and its environment interact in a highly non-linear way. The game as a whole does not evolve under a predefined state space, which is a subtle, yet crucial point. In mathematical models built from science, state space is almost always predefined[14]. Given the correct phase space to describe the system in question, the system can be tracked within a mapped space that is defined (even infinitely). However, in the case of League of Legends and many other complex systems represented by modern data, the state space is cannot be defined. There is no knowing the number of possible species evolution can produce or the number of all possible lives humans can live.

Furthermore, it is almost impossible know the entire state of the game at any given moment. **To know the state of the game requires one to know the state of every player and everyone at Riot, including their mind and thoughts.** This is simply not possible. It is only possible to know what data has been collected about the game over time, including the evolution of League's base code. Therefore, we have already been restricted to many sorts of observations we could make about the game. Knowledge about what players feel, such as when their in-game characters die, has been lost and yet could be a crucial observation to understanding how the game changes over time.

Figure 2 shows a schematic of the players' exploration of the game space, as constrained by Riot Games. The possible game state space is not self-defined because it depends on several unaccounted factors such as technology that is implemented, creativity, and even the economy. In this sense, the game as a whole (players, platform and Riot Games) is open-ended, since its evolution will never repeat[13,15,16].

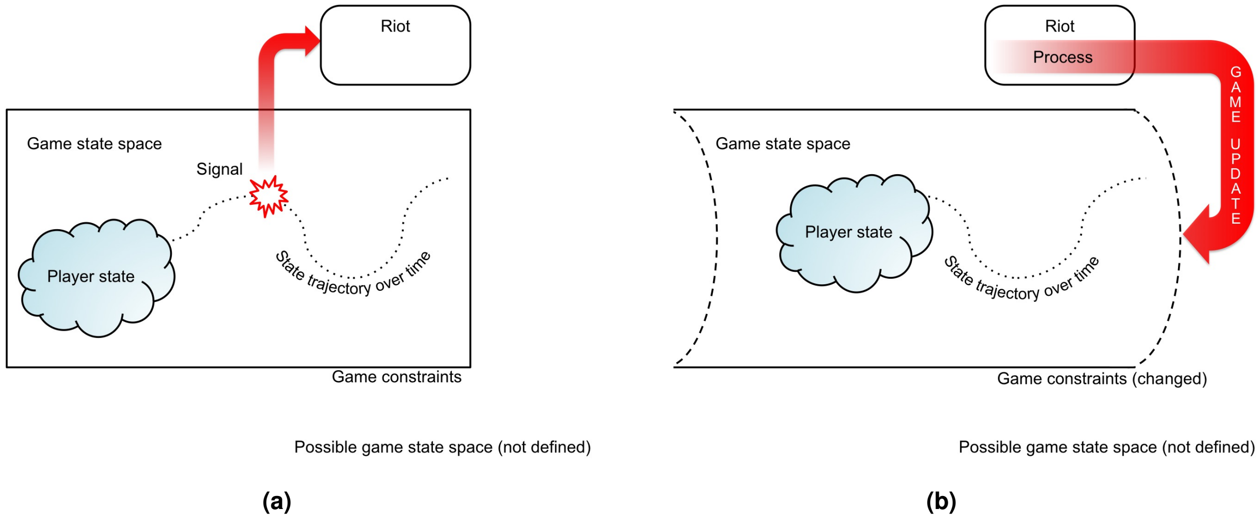


Figure 2. One possible abstract model to represent League of Legends as a whole. (a) In the possible game space, Riot's rule constraint a subspace for the players to explore. The players have a finite number of possible ways of creating data in the game state space. At some point during the players' trajectory, Riot Games collects some signal about the state of the game generated by the players. (b) Riot processes the signal in some way and changes the game constraints by updating the game. This allows new space for the players to explore.

By defining a constant constraint on the game space, the dynamics are guaranteed to repeat and open-ended evolution is not possible. Changing the constraints allows a system to generate dynamics that do not repeat, thus the player dynamics are also open-ended as long as the constraints are changing[16]. It is unclear how a signal is translated into an act on the game's current constraint, and Riot might not use a single function to determine it.

Collecting Data

Our question is this: Given current data that has been collected from the game, is it possible to predict (in some form or another) what *future data* will look like? It would not be efficient to process all of Riot's data for this analysis, so only a small subset is of the data is considered.

League is a very complicated and layered game in its own right. There are three maps, seven match modes, and eight tiers of skill. For this analysis, we only collected data from players that have the highest skill ranking on the North American server. League's skill ranking system is crowned with the Challenger tier: Only the top 200 players on each server occupy this tier of play. If a player is in the Challenger tier, they are most likely a professional player and stream their matches on Twitch (an online streaming service) for money. Several tournament players are in Challenger tier as well.

In League's player community, Challenger players often set gameplay trends. Since they are the best-of-the-best players, lower-level players often imitate their strategies and gameplay styles in hopes of climbing the skill ranks themselves. Hence, we assume that Challenger players are the players' main drivers. We note that there are other aspects of lower-skill gameplay that are not addressed in the current work. Although this game has many interacting complex parts like most complex systems, we emphasize that the main point of this study is not to determine if this system evolves in an open-ended way, nor does it explore what drives the open-endedness. Rather, the point is to determine how this system evolves mechanistically and if these mechanisms can be generalized to other levels of biological organization.

We also only consider matches that were played on League's most popular map, Summoner's Rift. Since we want to ensure these players are playing their best (because players can purposefully play bad games to entertain an audience), we only consider Ranked matches. Ranked matches determine if a player stays in Challenger tier or is demoted to a lower skill tier and are therefore taken more seriously.

Since even the abstract representation of the game showed in Figure 2 has many interacting parts, we focus our attention to only one part: The player state. Here, we consider two consecutive game "patches". Between the times that the game is patched, the game's constraints remain unchanged and constant. As a starting point, it will be easier to assess two separate time periods in the data where the game constraints remain unperturbed by developer intervention. These patches, named 7.2 and 7.3, were used 6 Jan 2017 - 7 Feb 2017 and 9 Feb 2017 - 22 Feb 2017, respectively. One day between the two patches was discarded since the game deployed patch 7.3 that day.

As a side note, all game data is made freely available to the public by Riot Games, and can be downloaded thorough Riot Games' server API[17]. To save time, we used a pre-written Python 3.X package called Cassiopeia[18] as an interface to connect to the API. Datreant[19] was also employed for fast and easy data sorting and exploration.

Data Represented as Networks

Players interact with each other in a large variety of ways in the game. The most common interaction is through individual game matches where one player wins a game against another, along with their respective teams. Interactions are most commonly modeled as networks, where agents are represented as nodes and their interactions are represented as edges. Here, it would follow that League's data could be represented as a network and using our current knowledge about networks, can be interpreted accordingly.

Dynamics Between Players

Riot's changes to the game are assumed to be changes to the game's constraints. Players (Challenger tier players) generate data by playing matches (ranked matches on Summoner's Rift). Between patches, the game is fixed in the sense that it has a static code underlying the game and a pre-defined state space; there are only so many different things players are able to do with the game's current code. In physics, systems typically evolve under a pre-defined state space. This makes them easier to understand conceptually. Equations are used to quantify how systems move through and explore the pre-defined state space. For this reason, this part of the data is the best place to start to eventually understand the game's dynamics as a whole. In other words, we begin our analysis by considering player-generated data between Riot Games' game changes, where the game's code is unchanged and the constraints that the players play under are constant.

How are the players moving through this space? In other words, what aspect of the game to we wish to study and ultimately want to predict? Presumably, players want to win matches and are picking strategies to help them achieve that goal. So is the players' trajectory their way of discovering optimal strategies to win the game? There is a subtle conceptual pitfall here based on what the data represents. Despite Riot Games' attempts to make all of its champions equally competitive, there is no reason to assume there is a pre-existing optimal champion or strategy to use. All knowledge about how the players explore the game state space (including Riot Games' knowledge) is generated solely by player data. There are only a finite number of possible states the data can be in at any given time. There is also no explicit "best" strategy that is optimal at any given time. However, players supposedly seek a type of optimum by exploring possible strategies and using the ones that win the most games. This could be an analogy for biology since organisms and species evolve by exploring the possibility space of phenotypes and use the ones that are the most "successful" at a particular time.

There are several different ways to quantify how the players explore possible strategies and evolve the current "metagame." The term "metagame" is akin to the most popular strategy being used at a given time. Physically, the game consists of people, electronic hardware, and software. Virtually, a metagame can be described in several different ways, depending on the question at hand. As an example, asking what the metagame is at a certain point in time is much like asking what state the United States' Republican party is in at a certain point in time. In an attempt to understand the game in alignment with Riot Games, we are going to use a metric that players use to pick their strategies for every match: What champion beats what other champions. Champion selection is a core mechanic of selecting a strategy. If a player sees the enemy team picks the champion "Teemo", the player is likely going to pick a champion that can beat "Teemo," such as "Talon." This decision is guided by knowledge from past experiences. In previous games, the player was successful in beating "Teemo" with "Talon" (or perhaps participated in games where they observed this interaction between other players) several times. In other words, players have some general knowledge about the metagame.

Since there is no explicit optimal strategy to use as a goalpost for the players, we can only use only player-generated data to determine which champion beats what. Let's say champions A, B, C, D, and E are used on Blue team, and champions V, W, X, Y, and Z are used on the Red team for a single match. Blue team wins the match. We say that in this match, champion A beat all Red team's champions (V, W, X, Y, and Z), W beat all Blue team's champions (A, B, C, D, and E), etc. This is a generalization of high-level play, since all players on Red team interact with all players on Blue team frequently during a single match. This also simplifies our analysis. This can be represented as a network, where if the character Teemo beat the champion Tristana, the interaction is represented by a directed edge from node "Teemo" into node "Tristana", as shown in 3.

For an aggregate of matches, edges have weights to represent exactly how badly "Teemo" beats "Tristana" during those matches. Edge weights are quantified in the next section.

Using Networks

To represent the players' internal dynamics under constant constraints (during a single patch, when the games' code is unchanged), we constructed a champion counter network from collected data to see how a champion's "power" changes over

Figure 3. "Teemo" is on the Blue team and "Tristana" is on the Red team. If Blue team won, we represent the "Teemo"/"Tristana" interaction on a graph like so.

time. In this context, power is used to describe how dominant a champion is over other champion, given a set of match data. Dominant or powerful champions are often more likely able to win games, regardless of the rest of the team's performance. Players often consider particular champions to be powerful during certain days or weeks, while other characters could be considered under-powered ("Who would ever play *that* champion?" a player might remark).

At any given minute throughout the day, anywhere from 10 to 100 matches are being played in this pool of players. Matches can last anywhere from 20 to 60 minutes (a hard minimum on 20 and a rough higher estimate on the 60). It is unclear if the game has a natural time scale, which makes it difficult to identify individual time steps. All matches were binned into days according to their start times. This is arbitrary, but since Riot Games' changes occur anywhere between 1 and 3 weeks, it seems like a reasonable time unit for network analysis.

Edge weights are constructed in the following way for a given day:

$$
\begin{aligned}
w_{i\to j} &= (\text{popularity of } i \text{ vs } j) \times (\text{winrate of } i \text{ beating } j) \\
w_{i\to j} &= \frac{\text{matches of } i \text{ vs } j}{\text{total matches}} \times \frac{\text{matches where } i \text{ beat } j}{\text{matches of } i \text{ vs } j} \\
w_{i\to j} &= \frac{N_{i\to j}}{M}
\end{aligned}
\tag{1}
$$

Since edge weights change on a daily basis, the network is dynamic with respect to time.

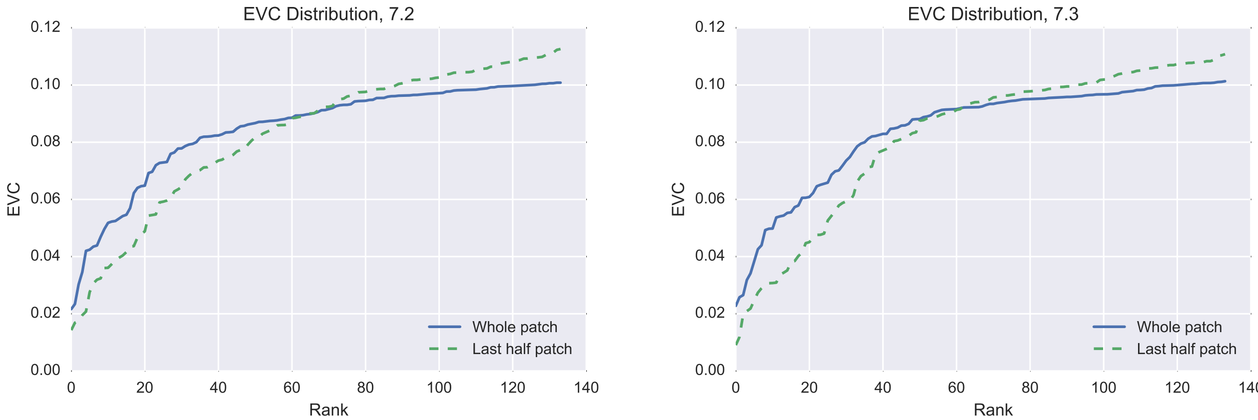
Analyzing Networks

For this analysis, we use eigenvector centrality (EVC) as a representation of a node's power. For a weighted adjacency matrix W with EVC values x_i, eigenvector centrality is defined as:

$$
\begin{aligned}
\mathbf{W}\mathbf{x} &= \lambda \mathbf{x} \\
\mathbf{x} &= \mathbf{W}\mathbf{x}\frac{1}{\lambda} \\
x_i = \mathrm{EVC}_i &= \frac{1}{\lambda}\sum_{j=1}^{n} w_{i\to j} x_j
\end{aligned}
\tag{2}
$$

Google uses a modified version of this to determine the webpage rank as the result of a Google search[20]. Colloquially, EVC represents how much influence a node has in a network. It not only considers how many in-degrees or out-degrees a node has, but the degree of nodes it is connected to as well. A node with 5 connections has a higher EVC value when the nodes it is connected to are also highly connected. With this metric, we are interested in two questions to understand how a champion power hierarchy is formed and maintained. (1) What happens to the power over time in champions that were changed in the last patch? (2) What happens to the power over time in champions that are going to be changed during the next patch?

To gain a general sense of the distribution of power in the game, all data for a single patch was aggregated into a single network, shown in 4. The distribution in 4 shows the EVC ranking of champions over an entire patch, for both patch 7.2 and 7.3. This represents a larger time scale than individual days.



(a) **(b)**

Figure 4. (a) Ranked EVC distribution of nodes over the whole patch 7.2. The blue line represents EVC measured over the entire patch's data aggregated onto a single network. The green dashed line shows the same, except for the last half of the patch. (b) Likewise for patch 7.3.

The same analysis was completed for individual days, but we found each individual day had an approximately linear distribution. Since the distribution over the whole patch is non-linear, this indicates that points on the linear daily distribution move up and down the distribution day by day (5). This figure shows the same ranking of EVC values, but on the time scale of individual days. This distribution is linear, as opposed to the non-linear distribution for the larger time scale.

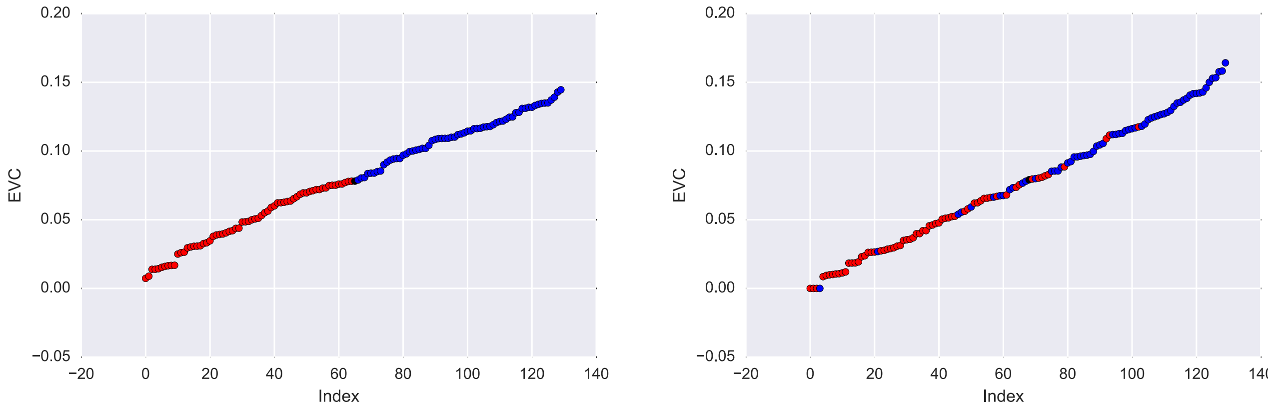
In 4, the majority of champions on the upper half of the distribution continue to remain on the upper half day by day. The non-linear distribution on a larger time scale and the linear distribution on a smaller time scale remained largely invariant for their respective time scales. However, the non-linear distribution from the larger time scale suggests there may be an underlying dynamic on how the linear daily distributions change point ranking. This phenomenon *may* indicate a type of dynamic equilibrium in the game's dynamics, which is characterized on a larger time scale by a non-linear distribution. The EVC values for individual nodes over time were analyzed to address this phenomenon. No clear trends were found for the accumulation of nodes (champions), so it was more fruitful to analyze specific champions, particularly ones that were changed during the two patches. But even after doing so, there was not a consistent translation between EVC changes in individual champions over time, and Riot's changes of those same champions (6).

From 6, it is unclear whether Riot Games' patches were beneficial to the "power-balance" of the corresponding champions according to this analysis. In other words, it is unclear whether Riot Games' patches accomplished the goal of adjusting the champions towards a more equal distribution of power. In 6a, "Akali" was buffed (made better in some way) and "Camille" was nerfed (made weaker). Their power levels after their changes seem to suggest the opposite. Only a few champions seemed to have benefited from their changes. 6a suggests a better indication of what champions need future changes for the majority of champions, though still misses the mark on champions like "Rengar" and "LeBlanc" since their power is decreasing over time, yet they are nerfed anyways.

Criticisms

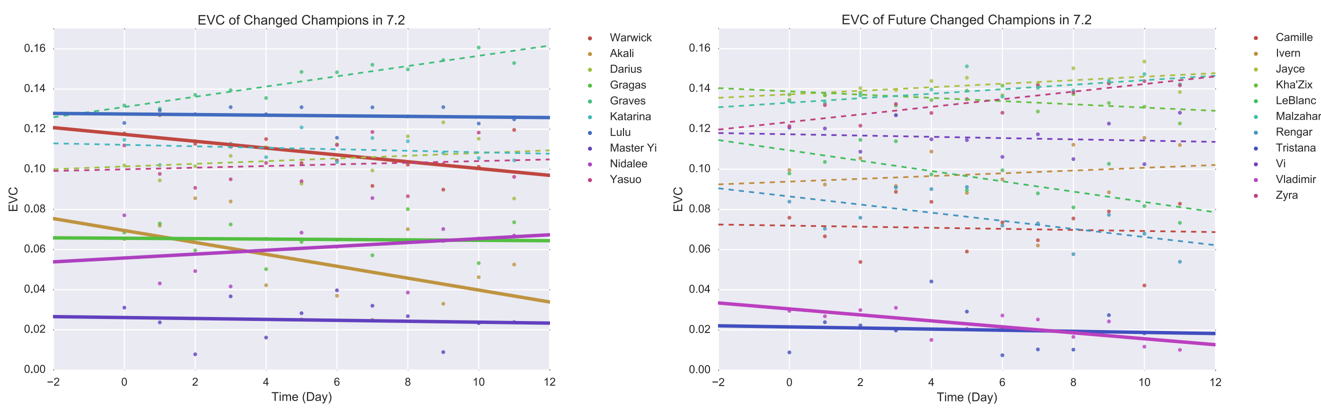
Since champion power levels depend on each other in a very intricate way, buffing or nerfing a particular champion doesn't seem to have the desired effect on that given champion. For example, if champion A beats champion B, nerfing champion A is an indirect buff to champion B. Since this system is highly connected, making changes to the game can perturb the network in unforeseen ways. A deeper understanding of how elements of the system (such as champions in this context) affect each other rather than analyzing the element dynamics themselves seems to be a more insightful direction of inquiry than a network approach.

Although networks are useful in understand the direct relationships between entities, an agent-based-modeling approach could also benefit us in understanding how these entities interact in a complex system. For example, we could model agents who randomly select champions based on a distribution derived from the data. Match outcomes could be decided on factors such as the agent's skill and data-based outcomes on which characters win against which other characters. From such a model, are the same EVC distributions produced over the rank of champions? We could also incorporate feedback, updating the champion-selection distribution based on the resulting EVC distribution. Such a model will certainly be worth exploring in the near future and would be a useful tool in understanding what factors affect these EVC distributions.



(a) **(b)**

Figure 5. (a) Patch 7.2 single day (the first day of the patch) distribution of EVC node values. Points on the upper and lower half of the distribution are colored blue and red, respectively. Colors simply correspond to their rank position on the first day of the patch. Without changing the colors of the nodes, the distribution of EVC is shown for the last day of the patch (b).



(a) **(b)**

Figure 6. (a) EVC values for champions that were changed in patch 7.2 over time. Solid lines represent champions that were “buffed” (made better in some way). Dashed lines represent champions that were “nerfed” (made weaker). (b) EVC values for champions in 7.2 that were eventually changed in patch 7.3 over time. These are champions that “signaled” Riot Games in some form or another.

On the other hand, the data provided already over-simplifies human players since it does not encompass their experiences as players. Thus, the networks we constructed to represent this data is a direct reflection of that. Constructing an agent-based model would accomplish much of the same task by omitting such data. While it is true that the model might reconstruct these distributions, this analysis might not yield a deeper understanding of the processes that govern living systems. Given there is so much missing data from our system of study, it is unclear how much these results reflect League of Legends.

Other Approaches

Aside from constructing a time-series of networks and applying measures to it, we could also address the state of individual players, given the data we have. While it is impossible to know a player's thoughts and feelings at any moment, we could assume a player can be represented by their history of gameplay. Although this method does not take into account how a player is influenced by YouTube videos or tutorials, we could assume a player remembers their past x matches and the outcomes of those matches.

Are players aware of the entire network's structure at any given time? Players only observe interactions between 10 nodes per match, so it is unlikely they observe interactions that include all the 140+ champions every day. Players observe a smaller network from only the matches they participated in. If their goal as a player is to win, are they increasing the EVC value on their champions' nodes in their smaller network of observed interactions? How do the EVC values of these smaller networks compare to the EVC values of the whole network?

In addition to these questions, it's important to highlight the two different kinds of League players. There are players who mainly stick to playing a single champion, and players who play whatever they think will help them win matches. They can be denoted as Type 1 and Type 2 players, respectively. Type 1 players who stick to a single champion over time have more influence on the properties of a single node. They actively maintain the state of a node and its connecting edges. In some sense, this is a lot like a system's stability. Type 2 players who only play certain champions to win and the player who are more exploratory and perhaps even more innovative. Their presence likely contributes to nodes that are likely more powerful. In another sense, they're a system's innovation-apparatus.

Using these assumptions and the given data, it could be useful to see how a player's history influences their future outcomes. For example, if a player has a match history over x matches that is an accurate reflection of the game history as a whole, is that player more likely to win? How does this change between Type 1 and Type 2 players? Is a team made entirely of Type 2 players more likely to win than a team of Type 1 players? And so on.

A Possible Shift in Theory

So what is to gain from this study on League of Legends? Mainly, it is not simple to analyze data that represents complex and/or living systems. This is largely because data is merely an abstract *representation* of a collection of real things, which is perhaps the reason why its analysis has not yet yielded to a comprehensive theory of living systems. Of course, this could be due to the fact that we simply need to do more data analysis– a *lot* more– although this approach comes with its own set of problems as well[21,22]. In short, given multiple large sets of numbers (multiple sets of data), correlations are always guaranteed to appear[21], even if they are, for all practical purposes, absurd[23].

In the rest of this paper, I suggest a new approach to representing the world so that better data (more encompassing data for real living systems) can be collected. This is by no means rigidly mathematical, nor founded on solid axioms. Instead, this is simply an exercise in simplifying what has been discussed so far in a way that is deliverable in less words and could serve as a starting point for future, more rigid mathematical discussions on living systems.

There seems to be a general agreement that living systems need at least three components to exist: (1) A physical substrate to exist in, like molecules and atoms, (2) a way to process information in some way through sensing, (3) and a meta-rule that governs how information processing should change over time, e.g. evolution. However, the exact relationship between these three ideas remains largely unclear, particularly how information processing and its meta-rule emerges from matter. Are the laws of nature somehow embedded within current mathematical models? This is likely not the case. Consider the quantum description of a hydrogen atom. It is highly unlikely that using the same description for the Krebs cycle, if even at all attainable, would be of any use for understanding the aerobic processes of a cell.

The purpose of this section is to explore the possible relationships between these three components in a way that does not exclude any of the aforementioned aspects of living systems that have been thus far ignored. I propose a general framework that could possibly describe their relationships, which is then related to other living systems, including our example dataset in League of Legends.

Physical Substrates

Coulomb's law does not exist where there are no charges[1]. Therefore it may be impossible to find the laws of life in systems that are not living. A single bacteria in a petri dish may not actually be a living system even though it is, in fact, alive. There should be a strict distinction in our awareness of "life" and "alive". Although a single molecule may not be considered alive, such as RNA, it is considered by most be an essential part of life. The current search for extraterrestrial life is focused mainly on the detection of biosignatures, which may not be alive themselves but are an indication that life is present. Without life, those biosignatures would not be present, whatever they may be. A mahogany table, which is also not considered alive, would be considered a biosignature of life since it is unlikely to be found where life is not present. Yet, biosignatures are significant to the environment of living systems (think of your own home). On a cellular level, if we are prepared to say an RNA is life although itself is *not* alive, then we must be prepared to accept entities produced and used by life as *a part* of life. That is to say, a mahogany table is certainly a part of a living system since it has been generated by life and is used by life.

It is therefore too difficult to delineate the boundary between life and its environment. This is my first preposition: that we should move away from drawing a boundary around life and its environment. There should only be physical pieces of reality, whatever those may be, whether or not we can sense them.

Information

For humans, reality is defined by what we sense. Even though there is a strong assertion to announce that objective reality exists, we can only attribute this to the agreement of common observations between people. This consistency, although useful for constructing theories about reality, is not reality itself because it only exists within observers.

Sense is the capturing of information. Information, in some way, can be colloquially understood as "a difference that makes a difference[24]". As far as this can be understood using the theories of computation and information, this includes necessary things like memory, state, and some sort of interactions between entities. Living systems have this ability to sense and to perceive information about internal or external things. Being able to collect this information greatly depends on the observer's ability to sense and perceive, which could be attributed to the current state of the observer.

Memory

Within physical states (or configurations/arrangements of matter), the notion of memory is embedded. In computational theory, memory is the preservation of past states through the current configuration of bits in a current state. A system has memory if it retains properties about its previous configurations in its current configuration. However, the movement from bits to the real physical world is difficult, since it moves from an abstract realm to one laden with specific details. In the physical world, bits are understood by human notions of yes or no, up or down, left or right. If I lay the popcorn bag this-side-down then it is distinguishable from laying the popcorn bag this-side-up. The world outside computers preserves our understanding of memory in the way that if a person changes something, then that change can be seen for some amount of time. If I change the facing of my popcorn bag, then that action is transcribed in the world by the bag's facing[2]. In this sense, a person's memory is no longer limited to their own brain. A human's physical interactions with the world ensures that memory is preserved in the world in some way.

As living systems change external properties about the world (the arrangement of things within a room, for example), cause-and-effect are greatly entwined with the idea of recovering the past from physically inscribed memory. Imagine a banana taped to a ceiling of a kitchen. How did it get there? Clearly, there must be some cause for the effect to having a banana affixed to the ceiling. The number of causes is also relatively large, even on a macroscopic scale; *who* put it where, rather than what arrangement of individual atoms put it there. This loss of information relevant to an individual observer has been postulated as the origin for the arrow of time[25]. This phenomenon is most commonly studied in thermodynamics and statistical mechanics: Given a current state $s \in S$, where we can assume S is finite, albeit unimaginably large if the universe is finite (otherwise we assume S is infinite), the number of possible causes $|C|$ from the point of view of an observer $o \in O$ is greater than 1.

$$|C|_o > 1 \;\mid\; s \tag{3}$$

Perception

Complexity theory is so far constrained to our own perception and measurements of a system that we ourselves are embedded in. Knowledge can only be constructed by our observations, which are constrained by our physical senses. As first-order observers, we may not be in a position to provide an observation of complex systems that has enough information to construct a consistent theory. If only we could become second-order observers and gain a bird's-eye view of a complex system then we might be able to gain more information to explain away what is currently unexplainable.

[1] This is a deep, deep philosophical argument, but here I am strictly asserting a claim on this side of the argument.

[2] This is also the purpose of keeping a notebook. By changing the properties of the physical world (writing in a notebook), the mind's thoughts are stored elsewhere than the mind itself. A person can access the notebook (physical memory) instead of preserving the thoughts within the mind.

Imagine an event that occurs within view of multiple observers. Do all observers perceive the same things? Unless each one of those observers is in exactly the same physical state (location and molecular configuration) then the answer is no. Yet we, as humans, assert there is consistency between us in various ways. The sky is blue, concrete hurts when fallen upon, and Nickleback is a terrible band. Even though we all receive different information about our world, we are able to use language that excludes minute differences to come to an agreement in perception.

This is the strength of the scientific method. It provides a set of requirements that multiple observations must meet in order to make a statement about the world. This is the basis that theories emerge. Newton's First Law is the language used to describe many many observations made under similar constraints. It is under this assumption that the word "theory" can be used to describe an agreement among observers about an event, or set of events, in a general sense. Within this context, observations refer to the ability to perceive information about something, internally or externally to an observer, which depends on the current state of the observer and the ability for the observer to perceive. In other words, perception can be an attribute of the state of an observer, which is a physical entity.

Observers

The observer is a physical entity made of physical components. The fact that it can "make observations" is attributed to the fact that it can be physically manipulated by external factors, much like a hard drive can perform read/write functions given a command. In this example, a hard drive is certainly an observer, since it perceives commands from a computer's user to change some aspect about itself. An observer could therefore be defined as a physical entity that undergoes differences when it receives a signal that prompts it to do so. However, this depends entirely on the signal as it does the current state of the observer. More on this later.

However, this implies that there is no differentiation between data and programs, or states and laws. And why should there be? Although it would be a positivist approach to assert that the laws of a system are the state of the system itself, this is not unlike many current theories in science. Coulomb's law does not exist where there are no charges, after all. Although this approach is insufficient to explain phenomena about the teleological nature of living organisms, it is at least consistent to appropriate the laws of nature as observed properties of nature itself. One approach in biology is that the structure of a molecule is most likely its function[26]; if a protein looks like a turbine, it most likely is a turbine. For macromolecules and smaller (at the very least), their physical attributes are enough to explain how they will interact with other entities of the exact same type. Even of interactions with entities of different types, such as ribosomes interacting with cholesterols, the rules of their interaction depends entirely on both entities' physical states without external perturbations. In the absence of external forces, both entities dictate how each entity will evolve to a future state.

However, it is completely absurd to discuss entities in vacuums, devoid of any interactions and messy, multi-leveled perturbations with the goal of understanding life. Rather, the world is full of observers on many levels of organization where elements of one observer may also be an element of another observer on the same or different scale. Thus the boundary around observers is not well-defined. For this reason, it is even more difficult to bound an observer from an environment. Instead, we must either retreat from the notion of observer, or understand the observer in terms of an emergent sense. In the latter case, it may in fact be true that observers have blurry or non-existent boundaries on many levels. For the former case, it is difficult to justify a world without observers. Although this description of the world may be useful, we as humans consider ourselves to be observers and it would be too difficult to traverse a path towards that agreement.

Emergence

To summarize the discussion so far:

- Living systems are composed of physical things and entities,
- Memory, the ability to sense/perceive, and information are attributes of physical configurations (states) of physical matter,
- Observers are composed of physical substrates and have the ability to sense, and
- The boundary defining an observers is not well-defined on many levels of scale and organization, along with the boundary between an observer and its environment, and

Under the existence of physical entities, information can be understood as the physical configurations of those entities and how those entities interact. Yet, we are still missing the third component of living systems: How do information-processing rules change over time? In other words, how does emergence in the most general sense (emergence of levels of organization, functions, processes, etc) happen in a physical system?

Since we are no longer assuming any explicit boundaries between entities and their environments, as well as boundaries between multiple observers, we are left with entities themselves. For completeness, all entities should be defined as sets of entities:

$$e = \{e\} \tag{4}$$

and all the states of entities are laws that dictate their evolution:

$$s_e = f_e. \tag{5}$$

The interaction between two entities (e_i and e_j) is defined by the combination of both their functions:

$$e_i \leftrightarrow e_j \quad := \quad f_{e_i+e_j} = (f_{e_i} + f_{e_j}). \tag{6}$$

Since an observer o is also an entity, we can define it as

$$o_i \quad := \quad e_j \;\mid\; f_{e_i+e_j}(s_{e_i}) = s'_{e_i} \neq s_{e_i}, \text{ where } i \text{ can equal } j, \tag{7}$$

which reads "an observer is an entity that, given an interaction with another entity, changes its state." The existence of an observer becomes a question of the property of an entity's state: an entity's state must be posed to be altered by the state of another entity in some way via an interaction. I was explicit in making sure an observer is allowed to interact with itself, which intuitively includes things like the observer as a whole being an internal process, although this is certainly up for debate. In addition, an entity $e = e_1 \cup e_2 \cup ...e_n$ can interact with the entities it is composed of, such as e_1. This framework allows for any entity, made of entities, to interact with any other entity, made of entities.

For fun, let's explore the limits of such entities:

$$\lim_{|e| \to \infty} \{e\} = \text{Universe}, \tag{8}$$

that is to say, the largest possible entity is the Universe. Furthermore, the state of the Universe dictates how it changes. In the other direction:

$$\lim_{|e| \to 0} \{e\} = 0, \tag{9}$$

meaning an entity composed of no entities is non-existent, an empty set.

Although this exercise has been in no means rigorous, it serves as a mental roadmap for thinking about the world in a way that allows the emergence of living things. Notions such as observers, levels of organization, boundaries, individuals, etc could be expressed using a self-referential framework. It could be worth pursuing the development of such a model to see if emergent functions like experience, meaning, and cooperation are possible. However, it still remains unclear that these properties of everyday life could be captured in such a model because of its tractability. How is it possible to model an entity on a mammalian level with this framework? Given our current implementation of programs and data in the context of computers, this may not be possible to explore using modern technology. Because modern computers are rooted in the idea of applying programs to sets of data rather than fusing programs and data together within the same physical state, living systems may quite simply impossible to understand under our current computational framework.

Wrap-Up

So far, we've only considered a few components of League's dynamics given the data currently available to us. We could do many more things with the data, including using perturbation centrality measures [18] on our time series of networks since game changes occur frequently. We could also explore other ways of quantifying the players' dynamics besides using a network, such as the agent-based models discussed previously. In addition, the idea of smaller player-observed networks described in the last section can be studied for practical purposes. Do players have unique "personal network" unique to their own match history or is the network topology similar for any given Type 1 player? It could be possible that if players have personal networks, cheaters can be detected. The most common form of cheating is when players purchase high skill-level accounts to use them as their own. Another is using external code to artificially improve a player's skill. Cheating could be flagged by tracking sudden topological changes in a player's personal network. However, as discussed at length, the results of these analyses might not be useful in the end for understanding living systems, since the data is already an abstract representation of League.

Understanding the rest of Figure 2 is beyond the scope of this paper, but can be addressed in future studies to perhaps collect better data. In particular, what signal is Riot Games gathering from the players' dynamics? How is this signal processed and translated into game updates? How do the game's constraints on the players evolve? Finally, how do all these dynamics interact and aggregate to evolve the entire game as a whole? It could be very useful to spend some time at Riot Games and observe how the game developers operate firsthand. Although it is impossible to observe every League player to an exact detail, it is at least possible to observe Riot Games employees firsthand. Already, they have done extensive internal research on player behavior from the in-game chat logs[27]. They currently use chat log information to implement rules on players who excessively curse or bully, with hopes of making the game an even more positive experience.

Physical laws as we currently understand them are insufficient to describe biological phenomenon such as heredity, adaptability, and the number of global tweets per minute. Perhaps the reason this quest is so difficult is our lack of a fundamental understanding of biological data, as discussed at length. While the solutions here are represented as a general framework for future discussion, it is still unclear how to implement those ideas tractably. Much more discussion is needed on the matter.

References

1. Whitehead, A. N. Symbolism, its meaning and effect. Barbour-Page Lectures (1927).
2. Whitehead, A. N. Process and reality, an essay in cosmology. Gifforf Lectures Delivered in the University of Edinburgh (1928).
3. Whitehead, A. N. *Nature and Life* (Cabridge University Press, 1934).
4. Big data. URL `ibm.com/software/data/bigdata/`.
5. DeDeo, S. Collective phenomena and non-finite state computation in a human social system. *PLoS ONE* **9**, 109–122 (2011).
6. Chalmers, D. W., Francis, C. C., Pepper, N. M. & Bedau, M. A. High-content words in patent records reflect key innovations in the evolution of technology. In *Alife XII Conference*, 838–845. Artificial Life 12 (MIT Press, Cambridge, MA, 2010).
7. Buchanan, A., Packard, N. H. & Bedau, M. A. Measuring the evolution of the drivers of technological innovation in the patent record. *Artif. Life* **17**, 109–122 (2011).
8. Oka, M. & Ikegami, T. Exploring default mode and information flow on the web. *PLoS ONE* **8** (2013).
9. Oka, M., Hashimoto, Y. & Ikegami, T. Open-ended evolution in a web system. In *Late Breaking Papers at the European Conference on Artificial Life* (2015).
10. Hansen, D. How to become a data scientist in 2017. URL `https://www.forbes.com/sites/drewhansen/2016/10/21/become-data-scientist/#947bbe187d33`.
11. Riot manifesto (2017). URL `riotgames.com/riot-manifesto`.
12. Taylor, T. *et al.* Open-ended evolution: Perspectives from the oee1 workshop in york. *Artif. Life* **22**, 408–423 (2016).
13. Banzhaf, W. *et al.* Defining and simulating open-ended novelty: requirements, guidelines, and challenges. *Theory Biosci.* **135**, 1–31 (2016). URL `http://dx.doi.org/10.1007/s12064-016-0229-7`. DOI 10.1007/s12064-016-0229-7.
14. Longo, G. & Perret, N. *Building Theories*, chap. Rhythms, Retention and Protention: Philosophical Reflections on a Geometry for Biological Time (Springer, 2017).
15. Ruiz-Mirazo, K., Peretó, J. & Moreno, A. A universal definition of life: autonomy and open-ended evolution. *Orig. Life Evol. Biosphere* **34**, 323–346 (2004). URL `http://dx.doi.org/10.1023/B:ORIG.0000016440.53346.dc`. DOI 10.1023/B:ORIG.0000016440.53346.dc.
16. Adams, A. M., Zenil, H., Davies, P. C. W. & Walker, S. I. Formal definitions of unbounded evolution and innovation reveal universal mechanisms for open-ended evolution in dynamical systems (2016). ArXiv: 1607.01750.
17. Stats straight from the source (2017). URL `developer.riotgames.com/`.
18. Cassiopeia (2017). URL `github.com/meraki-analytics/cassiopeia`.
19. Datreant: Persistent, pythonic trees for heterogeneous data (2017). URL `http://datreant.readthedocs.io/en/latest/`.
20. Austin, D. How google finds your needle in the web's haystack (2017). URL `http://www.ams.org/samplings/feature-column/fcarc-pagerank`.
21. Calude, C. S. & Longo, G. The deluge of spurious correlations in big data. *Foundations Sci.* 1–18 (2016).
22. Buiatti, M. & Longo, G. Randomness and multi-level interactions in biology. *Theory Biosci.* **132**, 139–158 (2013).
23. Vigen, T. Spurious correlations. URL `http://www.tylervigen.com/spurious-correlations`.

24. Brender, N. M. Sense-making and symmetry-breaking: Merleau-ponty, cognitive science, and dynamic systems theory (2012). 2012 CSCP Graduate Essay Prize Winner.

25. Maccone, L. Quantum solution to the arrow-of-time dilemma. *Phys. Rev. Lett.* **103** (2009).

26. Abbot, E. S. The causal relations between structure and function in biology. *The Am. J. Psychol.* **27**, 245–250 (1916).

27. Hsu, J. Inside the largest virtual psychology lab in the world (2015). URL https://backchannel.com/inside-the-largest-virtual-psychology-lab-in-the-world-7c0d2c43cda5.

Acknowledgements

Special thanks to Riot Games' API team for allowing all League of Legends data to be public and easily accessible. Here's a shout-out to David Dotson, the creator of Datreant. This project would not have been possible without his help and tutorials. In addition, many thanks to my extremely supportive advisor Sara I. Walker, my professors Xin Wei Sha and Manfred Laubichler, and my colleague Pedro Nardelli for several stimulating and fruitful discussions.

Additional information

This project along with several others on League of Legends can be found on www.nostrateemo.com. All developments, code, and future updates can be found at this site.