2019

A Framework for Exploring Metadata

IN TRADING CARD GAMES

MCLEOD, JERICHO

Table of Contents

Background and Motivation

The trading card game, Magic: The Gathering™ (MTG), produced by Wizards of the Coast, LLC, has become one of the largest such games in the world since its inception in 1993. Over that time, it has grown to have 35,000,000 active players (Webb, 2018) who build and compete with 60 cards decks chosen from a set of 19,989 cards (Gamepedia, 2019). The quantity of cards and deck size mean there are 1.226859x10^176 possible combinations in total, not accounting for decks that are non-functional, and not considering sideboards, a second set of cards from which a player can reconstruct a deck mid-tournament. Given the complexity of the game, the rate at which new cards are added, and the element of randomness inherent in games involving shuffled decks of cards, it is unlikely to ever be solved.

Despite the complexity of the sample space players tend to converge on a subset of cards with preferable properties. This reduces the computational expense of studying the format a great deal while leaving space for interesting questions to arise.

The prize pools of tournaments are high enough to encourage such changes in competitive decks despite the costs. Players choose decks based on cost, performance, and some other components this project hopes to move toward identifying. The costs of decks, and of swapping between two decks at a particular time, is estimable given the public market for cards. Performance can be measured to some degree through tournament results, though the random aspect of the game confounds this to some degree.

This project aims to build a framework for modeling transitions between decks as an early component in understanding behaviors in this type of sample space.

Related work

“The Network Picture of Labor Flow” initially introduced me to the concept of abstracting movement of employees (Lopez, Guerrero, & Axtell, 2015). This paper was viewed as a step in the process to creating a novel economic forecasting method, which was made possible by fact that the research was conducted in a relatively mature space with comprehensive data available.

This inspired me to consider less mature sample spaces and how to approach identifying salient features. Material introduced within CSI-703, specifically, “Perception in Visualization,” led me to consider visual features as a way to identify desired outliers or patterns quickly in a network (Healey, 2017).

Research Purpose

Studying behaviors computationally in a large sample space is costly. To reduce costs, ideally exploratory analysis is conducted that identifies the key components that help understand which features of a sample space are interesting and bear further study. This project aims to build a useful tool for identifying these features in the sample space of Magic: The Gathering's legacy tournament results by visualizing it as a network. Decks with non-zero intersections will have edges that ultimately will bear some cost, while the decks themselves will be the nodes. Visualizing this space as a network allows a researcher to more easily understand a specific player's available options when making decisions in terms of which decks to play.

The next phase of this project will be to add tournament results to nodes. This would allow a researcher to visualize more completely the information that may have contributed to player decisions. An added benefit to publishing such a tool is information transparency for players; by viewing their current deck in this tool, a player could easily determine the cards necessary to transition to other decks which may be more competitive or interesting.

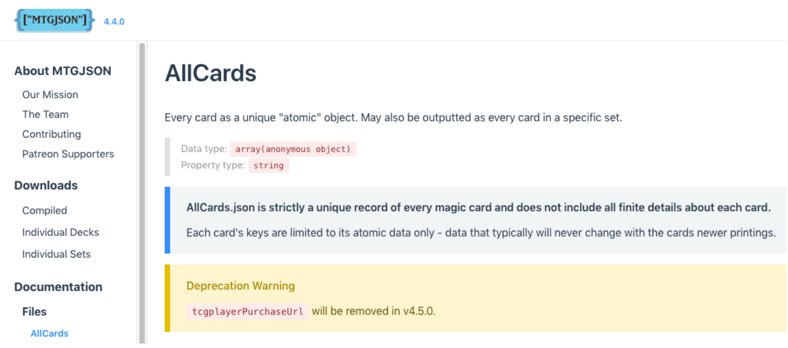
The final stage in creating a framework to model the Legacy deck sample space would be to add a temporal component. This would allow a researcher to visualize how changes in costs and tournament performance altered flows across time as shown by the number of players utilizing specific decks.

Once completed, this tool will make used to identify interesting flows using visual contrast; that is, those that are not explained by some function of cost and performance should carry some visual encoding. This project is ultimately a tool to elucidate further questions by allowing the examination of expected patterns in data and contrasting expectations with observations that is useful in other similarly situated sample spaces.

Data

Data Sources

Data for tournament results is published in numerous locations, but as planned, this data was obtained from the website MTGTop8.com. Additional card data was obtained from MTGJSON.com. Due to a recent update at the latter source, pricing data is also available there. Previously, pricing data was obtained via API with TCGPlayer.com, however, the need for momentary granularity over weekly in this project is minimal, thus API data collection was deprecated.



AllCards

Every card as a unique "atomic" object. May also be outputted as every card in a specific set.

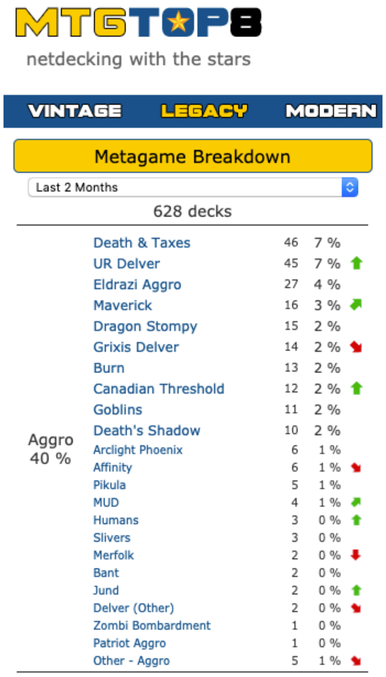
Data type: `array(anonymous object)`
Property type: `string`

AllCards.json is strictly a unique record of every magic card and does not include all finite details about each card.

Each card's keys are limited to its atomic data only - data that typically will never change with the cards newer printings.

Deprecation Warning

`tcgplayerPurchaseUrl` will be removed in v4.5.0.



Metagame Breakdown

Last 2 Months

628 decks

Aggro 40 %	Death & Taxes	46	7 %	
	UR Delver	45	7 %	↑
	Eldrazi Aggro	27	4 %	
	Maverick	16	3 %	↗
	Dragon Stompy	15	2 %	
	Grixis Delver	14	2 %	↘
	Burn	13	2 %	
	Canadian Threshold	12	2 %	↑
	Goblins	11	2 %	
	Death's Shadow	10	2 %	
	Arclight Phoenix	6	1 %	
	Affinity	6	1 %	↘
	Pikula	5	1 %	
	MUD	4	1 %	↗
	Humans	3	0 %	↑
	Slivers	3	0 %	
	Merfolk	2	0 %	↓
	Bant	2	0 %	
	Jund	2	0 %	↑
	Delver (Other)	2	0 %	↘
	Zombi Bombardment	1	0 %	
	Patriot Aggro	1	0 %	
	Other - Aggro	5	1 %	↘

Data Description

All Legacy tournament entrants available were gathered for the period January to March of 2019. This entailed a total of 611 decks, but only 43 unique named decks. Although initially unexpected, this project now has complete coverage of all reported decks for the studied period of time. Decks were averaged into their respected deck types; i.e., the 43 named decks represent the entire visualization. Within it, however, exist partial card quantities due to averaging and small variations in deck lists from player to player.

Data Cleaning

Minimal data veracity verification was conducted using Microsoft Excel for the speed of access. Additional JSON files for scraped data was created using purpose-built Python scripts. Link values use Jaccard Similarity; this data was calculated in Python, as was the average value of cards in each deck per named deck type. At the implementation level there is a large amount of data structure manipulation but minimal calculation in order to minimize response times to end users.

Exploratory Analysis

An approximate project plan was constructed based on a priori knowledge of the data that included visualizing the data using graphs. Once data was obtained, network graphs were created using MicroStrategy's Workstation. As shown below in Figure 1, the dataset was highly clustered. Filtering to only showing edges for sets of 4 cards made the graph somewhat intelligible (Figure 2). Further filtering to only one deck archetype (Control, Aggro, and Combo are typical divisions by the community of game participants) made the graph much more easily intelligible (Figure 3).

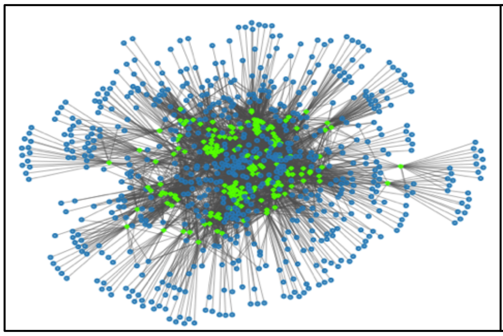


Figure 1

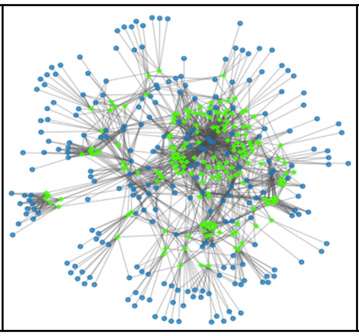


Figure 2

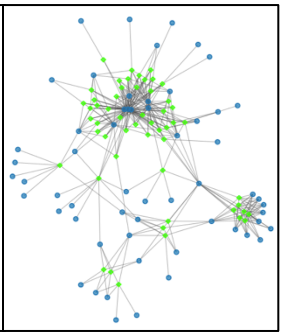


Figure 1

Decks are shown in green. Cards are shown in blue.

However, the tool used was relatively limited in functionality for data manipulation; only metrics included in the original dataset were available for filtering. It was decided that more useful metrics for filtering would be based on card types and costs. Card types were chosen for the high degree of variability between types within the format. Costs were chosen due to interest in patterns including cost in future research.

Design Evolution

Initially I planned to utilize force-directed graphs to provide interactivity with the visualization via the JavaScript library for data driven documents, or D3.js. This required additional time to gain some basic level of understanding of the language. However, upon creating such a visualization, several problems were encountered. The graph was expected to be dense when unfiltered, but it remained extremely dense even when filters were applied (Figure 4). The force of the graph was altered on a filter-by-filter basis, which allowed the various graphs to be expanded, but caused unlinked clusters to expand well beyond the canvas onto which they were rendered. This method also caused issues when portions of the graph were much denser than others (Figure 5). A border was then enforced along the boundaries of the canvas, but this caused smaller clusters to be flattened at the edges of the canvas (Figure 6).

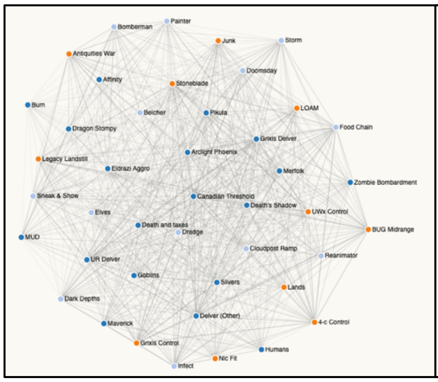


Figure 4

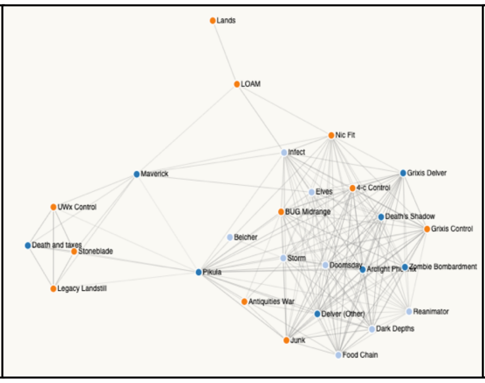


Figure 5

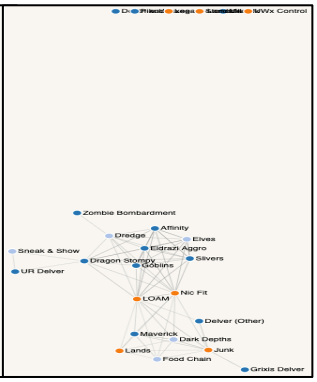


Figure 6

Though the graph was interactive, moving nodes presented a high degree of jitter in the animation. For these reasons, this format was discarded.

An alternative was needed that would fix layout issues, not re-render in its entirety, but still allow connections between decks to encode data. The best mechanism to visualize decks and overlap between components remained graphs, so a radial layout was chosen in order to fix positions. However, given the difficulties I had in implementing previously generated tools for creating a graph, this form was created from a blank slate.

The first component to be generated was merely the placement of nodes along a circle with edges drawn between them, shown below in Figure 7.

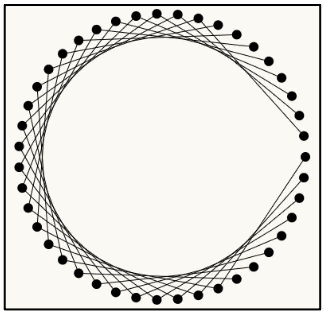


Figure 7

However, when all data was applied, the nearly-spectral nature of the dataset generated the image shown in Figure 8. It was clear that some level of interaction would be necessary in addition to filtering if all data would be shown in a single graph.

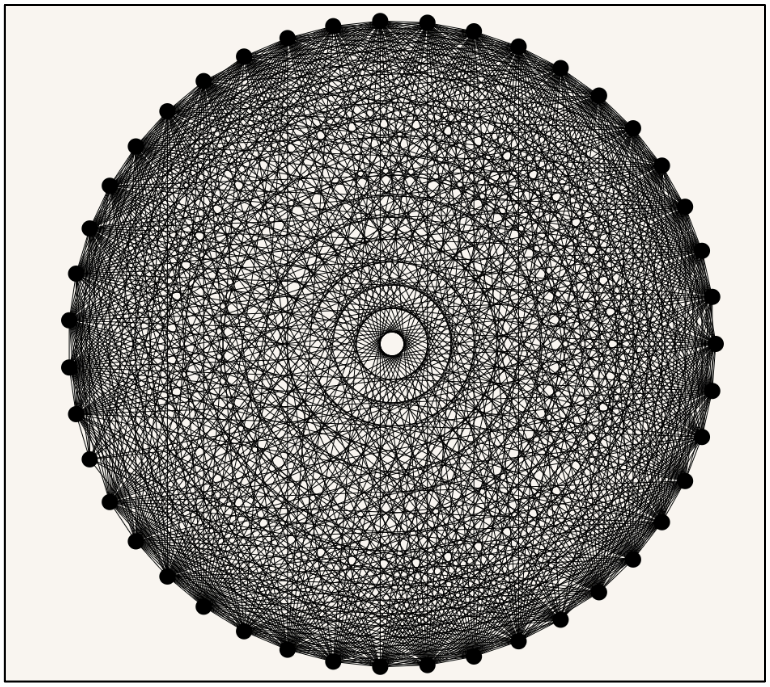


Figure 8

Highlighting a node of interest was added at this point so that its neighbors could be easily identified. In the initial graph this is nearly all nodes, as expected, but the addition of mouseover events to identify the similarity of the highlighted node and the neighbor being examined allows the relationships to be reviewed. Further, similarity is encoded in link shading, so that very-similar decks are visually identifiable the initial view or from a highlighted view.

In addition to highlighting, filtering was necessary to bring some level of understanding of the connections between nodes. After experimenting with various data points, it was decided to filter across buckets of costs and across card types.

Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

Encoded Similarity

Similarity between decks was calculated using Jaccard Similarity:

$$Sim(i, j) = \frac{|i \cap j|}{|i \cup j|}$$

Where *i* and *j* are nodes in the network, and $0 < Sim(i,j) < 1$. The similarity value is then used to determine the color of edge $a_{i,j}$ by scaling the value to a range of 0 to 255, then using it in triplicate to create a grey RGB value. This can be seen below in Figure 9 where some edges are noticeably darker than others. This serves to highlight visually which decks are more similar at a glance.

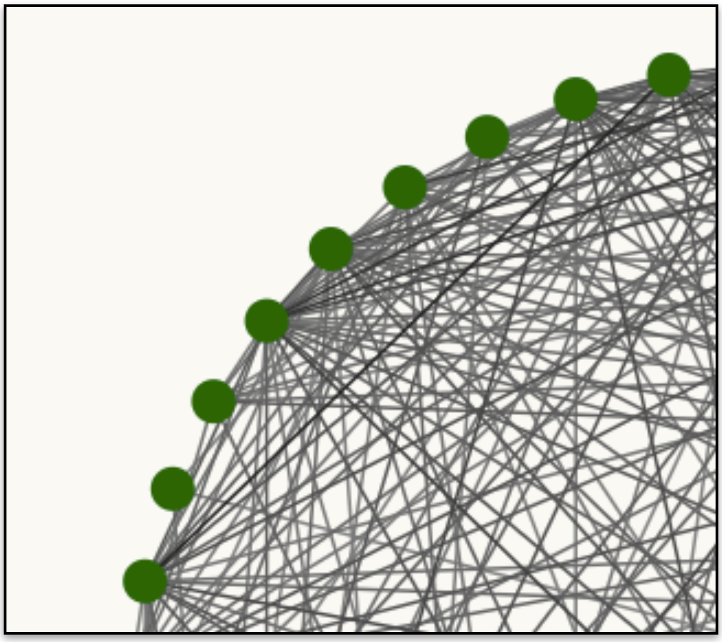


Figure 9

Filtering

The dataset is pre-filtered by user selections with price thresholds and card types. For price thresholds, edges are calculated only considering cards that met the threshold. Similarly, filtering by card type only considers cards of that type for calculating similarity between decks. Jaccard Similarity, the measure used to show commonality between nodes, is calculated as the intersection of two sets divided by union of two sets; the sets, or deck lists, are filtered, rather than merely the intersection. Notably, for otherwise unfiltered data, edges are only rendered if the number of common cards is at least 8; for filters by card type this is reduced to 4. The filter menu for pricing and card types, as well as the drop-down menu items, are shown below in Figures 10 and 11 respectively.

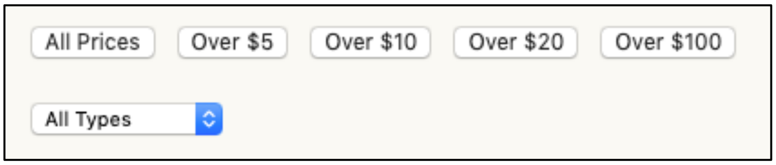


Figure 10

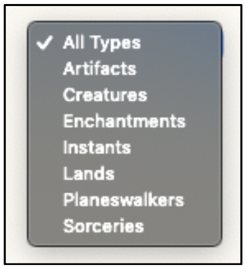


Figure 11

An example of filtering can be seen here: In Figure 12 all edges representing at least 8 common cards is shown, while in figure 13 only those with at least 8 common cards costing at least $20 are shown.

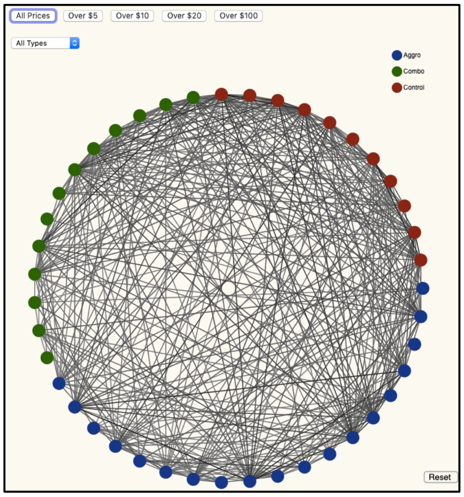


Figure 12

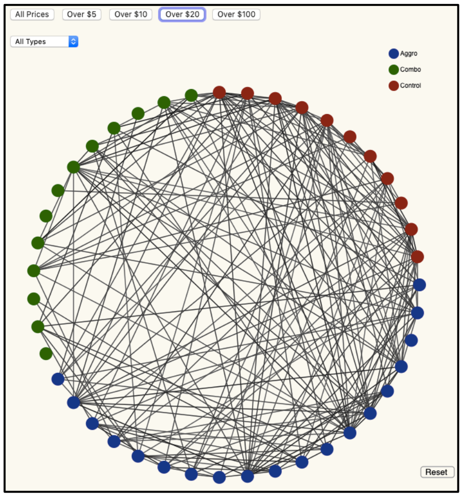


Figure 13

Legend

Deck archetypes were encoded as node color; this is communicated to users in a simple legend chart shown in Figure 14. Surprisingly, as basic of a feature as this encoding appears to be, it allows a great deal of insight to be drawn about common factors across and within archetypes using this visualization.

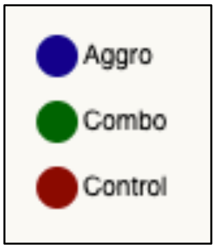


Figure 14

Highlighting

When a node is selected, its edges are darkened. Nodes that are not neighbors to the selected node are shifted to a lighter color, and all edges not connected to the selected node are shifted to a fixed shade of light grey. The before and after of this can be seen in Figures 15 and 16. This feature allows users to easily determine a node's neighbors in a visualization that is at times very dense.

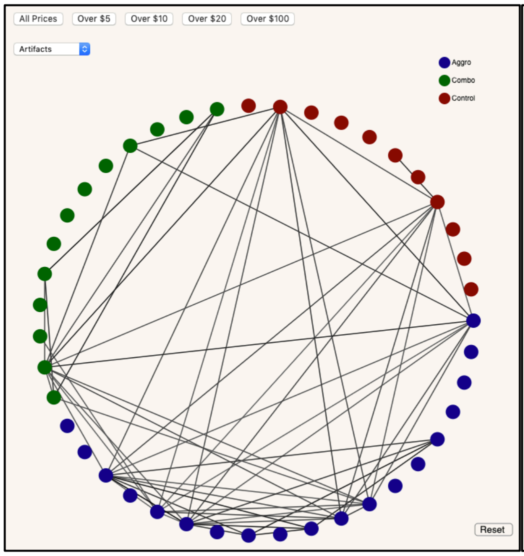


Figure 15

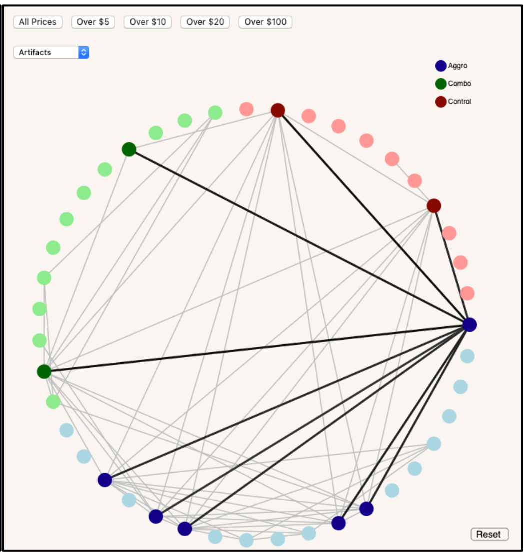


Figure 16

Tooltips

Two different types of tooltips were utilized in this visualization. Hovering over a node displays the Deck Name of the node, as shown in Figure 17. These render to the right of the node and contain a background to preserve clarity when rendered on top of existing edges.

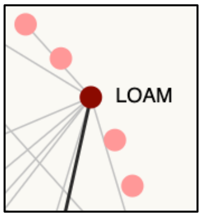
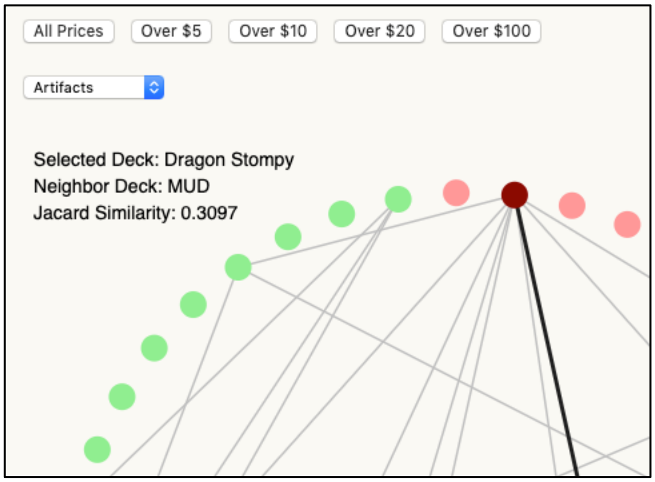


Figure 17

The second tooltip is statically placed at the top left of the visualization. This tooltip only appears when a node is highlighted, and the user has hovered the cursor over a highlighted edge. It displays the selected node name, the neighbor connected to it by the edge receiving the mouseover event, and the calculated similarity in the selected data set. This was initially drawn with the mouseover event coordinates but given the density of the graph and the narrow space available over edges, this offered limited visibility and usability. By placing it in a fixed location it does not interfere with further selections, allowing the visible time to not be limited to while the cursor is over an edge. This is shown in Figure 18 as text below the drop-down selector and above the green nodes in the graph.

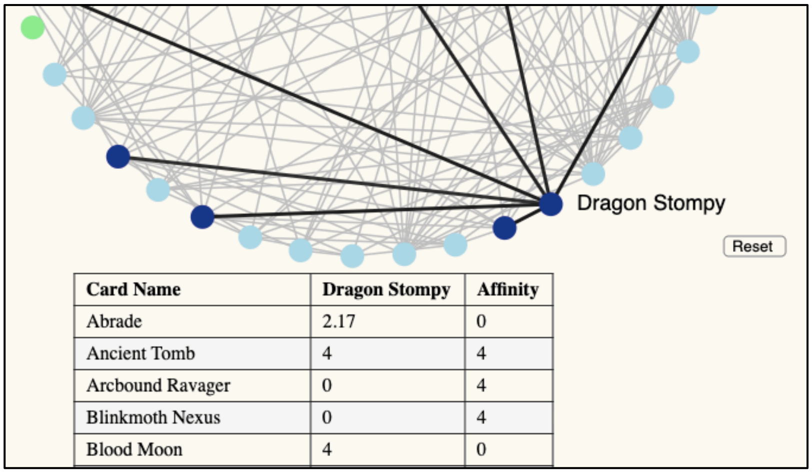


Selected Deck: Dragon Stompy
Neighbor Deck: MUD
Jacard Similarity: 0.3097

Figure 18

Data Table

When a node is selected, the deck list for that deck is displayed in an HTML table. When a second node is selected, the table is re-generated using a new list that is the combination of both nodes' deck lists and the values for each deck appropriately listed. An example of this is shown in Figure 19.



Card Name	Dragon Stompy	Affinity
Abrade	2.17	0
Ancient Tomb	4	4
Arcbound Ravager	0	4
Blinkmoth Nexus	0	4
Blood Moon	4	0

Figure 19

Evaluation

Data Improvements

Jaccard Similarity was utilized to show deck similarity. While this feature still exists in the final product, it is not being highlighted. The overall similarity of decks is less notable than card costs. In future versions that include cost, edge color will be on a gradient of binned cost categories or some relationship of cost and similarity that range from a light color to a nearly black shade of the same color for user friendliness. The overall cost of movement within the network is of more interest moving forward than raw card quantity similarities, while preserving similarity in such a measure would provide some average per card movement costs; both of these may be of interest.

Tournament results need to be integrated as well. Providing a measure of cost to success as a function of edge costs alongside expected tournament performance differences across edges moves further toward the goal of studying player behavior and also provides additional information to players in a novel format that is not currently available.

In order to make use of this concept as a framework for understanding decisions, temporal data need to be captured for pricing, deck lists, and tournament results. This allows computational discovery of how various elements influence one another.

Finally, decks need to be customizable. Allowing user input would allow additional deck data to be collected while also increasing the accuracy of the data provided to players in terms of transition costs.

Visualization Improvements

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it? Creating this project was challenging for a number of reasons. I lacked a strong background in the libraries recommended for visualizations initially but was also dissatisfied with existing software for the type of visualization I wanted to create. This raised the time investment necessary for prototyping visualizations and decreased the total iterative updates possible within the available time frame. While a visualization has been created for this project, there are many shortcomings that are often interrelated.

Transition

Most glaringly, the graph re-renders completely rather than transitioning smoothly to new states. This occasionally causes blank frames to render, which may inhibit viewer perception of changes due to change blindness (Simons, 2000).

Layout

The current layout shows neighbors with straight lights. Arcs would be better for not occluding lines and improving clarity in this layout. Additionally, a hierarchical layout on node selection would further improve the salience of deck similarity. Ideally, selecting a node should center it in the visualization with an animated, but reversible, transition. Neighbors should be shown in a ring around the selected node. Nodes two degrees separated from the highlighted node would then make up the second ring, repeated to N nodes and N rings. Ideally this radial layout transition would have node movements in arcs rather than straight lines (Yee, Fisher, Dhamija, & Hearst, 2001).

History

At present only steps k_t and k_{t-1} are captured, and k_{t-1} only in the table showing deck lists. A more intuitive implementation would include a history of steps with associated information such as costs of moves. If a transition-based visualization is to be used in future research, allowing users to move through them at will is necessary for them to be useful.

Another dimension of history in this visualization is examining prior time periods. This needs to be a user-controlled feature of the graph to avoid risk of losing relevance.

Bibliography

Gamepedia. (2019, 2 16). *Magic: The Gathering Statistics and Trivia*. Retrieved from Gamepedia: https://mtg.gamepedia.com/Magic:_The_Gathering/statistics_and_trivia

Healey, C. G. (2017, Oct 2). *Perception in Visualization*. Retrieved from NC State University Computer Science: https://www.csc2.ncsu.edu/faculty/healey/PP/index.html

Lopez, L., Guerrero, O., & Axtell, R. (2015, July 1). Retrieved from arXiv.org: https://arxiv.org/pdf/1507.00248.pdf

MTGJSON. (2019, 2 26). *MTGJSON*. Retrieved from https://mtgjson.com/json/AllCards.json

MTGTop8. (2019, 3 23). *Legacy Events*. Retrieved from MTGTop8: https://www.mtgtop8.com/format?f=LE

Simons, D. J. (2000). Current Approaches to Change Blindness. *Visual Cognition*, 1-15.

TCGPlayer. (2019, 2 26). *TCGPlayer APIs*. Retrieved from TCGPlayer: http://developer.tcgplayer.com/

Webb, K. (2018, 12 18). *With more than 35 million players worldwide, Magic the Gathering is giving back to its community with a brand new game and $10 million in esports prize money*. Retrieved from Business Insider: https://www.businessinsider.com/magic-the-gathering-announces-10-million-esports-program-for-2019-2018-12

Wizards of the Coast, LLC. (2019, 2 16). *Rules | Magic: The Gathering*. Retrieved from Wizards: https://magic.wizards.com/en/game-info/gameplay/rules-and-formats/rules

Yee, K.-P., Fisher, D., Dhamija, R., & Hearst, M. (2001). Animated Exploration of Dynamic Graphs with Radial Layout. *IEEE Symposium on Visual Information*.