Character Space

Julia Zimmerman, Philip Nguyen

May 2021

1 Introduction

1.1 Big picture goals

We believe that science is not discrete from everything else, and that it belongs to everyone. People are great at categorizing things and finding analogies between them; this is part of what meaning is. However, by putting things into categories, we sometimes sever their connections with the wider world, and, as time passes into habit and tradition, there come to be silo'd topics that have little to do with each other, not just in practice, but in how anyone thinks about them, even their experts. Specialization and generalization are wonderful tools of our group cognition, but, as with everything else, far from perfect in application.

Within the length of our arms, we'd like to try to contribute to improvements by involving interdisciplinary feedback in our work, and by doing our best to mitigate any barrier to entry to reading our paper. Our goal is that readers from any field will be able to read and understand our paper without encountering a barrier of jargon or formulae, and that this will support cross-pollination of fertile ground.

1.2 What do we want to know?

We want to know what fictional characters are atomically comprised of, because we think this reflects underlying cultural and individual cognition about what it means to be human. From there, we'd like to know along what axes we can best depict characters, and what the way the characters in a work occupy that space tells us about the story being told. In this paper, we aim to make an initial exploration of this premise, and establish if there is enough "there there" to warrant future study.

In future work, we'd like to track how characters change in stories, and compare those arcs with those found in previous work [Rea+16]. We'd like to examine how the depiction of characters (real and fictional) changes over time, and how the features that define characters are used by people to interpret real events.

1.3 What do we claim?

We claim that (1) it's reasonable to look for the meaning of characters in a physical framework whose components are a fairly small number of personality traits and (2) the way characters in a story occupy "Character Space" tells us about the story, drawing on an analogy with vowel space. Underlying our reasoning for these claims are the suppositions that (3) any model that you create to reflect the world also co-produces that reality, (4) that stories matter (and are worth studying), and (5) that metaphor is a fundamental part of meaning.

That's the big picture motivation for this (much smaller) project, which aims to look in particular at how authors construct and make use of fictional characters to support their storytelling.

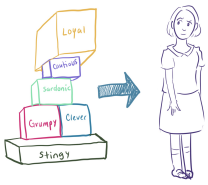

Figure 1: Personality traits add up to a fictional character.

1.4 Why do we think these underlying suppositions are reasonable?

1.4.1 Models

In order to represent a thing, your representation must be different than the thing[1]. Decisions and deviations are therefore unavoidable in any model. In general, a representation is a model is a story, which, beyond being seen as "cultural products [,] ingenious assemblies of conventions, patterns, and repeatable practices within historical and generic traditions" [Gol], must be seen as co-producing those traditions, patterns, and practices [Blo+20], and therefore reality itself (from the human perspective)[2]. In our case in particular, because we're discussing characters, tropes, and stories, there is also a tension between trying to accurately describe those things and a desire not to, by so describing, reinforce and repeat beliefs that we hope humanity can aspire to leave behind. Just by repeating e.g. stereotypes, without endorsing them, we contribute to their continued existence and reinforce them - the sort of mental heuristics people use mean we know this is inevitably the case. As Meg Murray says, "I can't think without feeling"[LEn07]! It's also the case that while we want to write a compelling paper, we don't want to be more persuasive than is justified[3].

1.4.2 Stories and Metaphor

There's a robust set of academic literature from various fields that supports the claim that stories matter [BC20; McA93; CM91; Cam08]: "Stories are an integral part of human culture. They allow us to express emotions, share knowledge, and to shape our perspective of the world (McKee, 2003)." Take, for example, the branches of ethics that deal with representational harms: representational harms can only exist if stories have a non-trivial impact on reality. There are at least 146 papers in NLP alone that touch on "bias", which implies the general acceptance of the claim that telling an unfair story can cause harm [Blo+20]. The existence of PR firms, backed by capital investment, implies that the market sees a value in the subtleties of a story, too. Other researchers have found that in attempting to bridge political and moral disagreements, sharing personal experiences are more effective than presenting facts in building mutual respect, an important precursor to productive conversation and social progress [Kub+21].

What this suggests is that stories can impact reality via the meaning we derive and prescribe to people, objects, and events, and by constructing relationships between them. How this manifests varies. For example, in math, we give soothingly precise definitions to objects and relationships between them, and from there we can confidently extract information that feels new and exciting despite it being somehow equivalent to what we started with. This seems to us like a fairly compelling reason to think that meaning, as we experience it, comes from perceiving connections (analogies/ metaphors) between things [LJ08]. Metaphors and analogies are of course core properties of stories, for example,

[1] ...or else it is the thing, not a representation of the thing.

[2] Taking this further, we're also concerned about our participation in the commodification of science. C.f. Liv Agar, the works of Adorno and Horkheimer, Natalie Wynn.

[3] "Rhetoric, declares Socrates in Plato's Gorgias, 'is the craftsman of persuasion which yields conviction but does not teach about the just or the unjust'. Conviction that bypasses the discrimination of justice is dangerous indeed."[Bee98]

in the form of tropes, which are shared across media when e.g. two works "'share' a property (for them both to exemplify, say, a particular shade of green) is for each to contain (instantiate, exemplify) a greenness-trope, where those greenness-tropes, although numerically distinct, nevertheless exactly resemble each other." [Mau13]

One of the most ubiquitous ways we encode meaning across individuals, time, and space, is via (broadly-defined) stories (via language) [CG16]. So, stories, packages of utterances[4] bundling together often-familiar analogies, are an elegant, efficient, and perhaps inevitable part of human cognition [Erw17].

Stories contextualize information and events in a relevant and meaningful way by connecting who we think we are with the happenings of the world; they form the basis of identity and empower us with a sense of personal agency [McA93; But90; FC96]. Furthermore, we think that the universe spanned by a story is meaningful to understanding the impact of the work and the perspective of the author(s) behind it. Works can be seen as reflecting the "inner maps" of authors [Gra73], and influencing the inner maps, empathy, and judgment of the readers. If stories are unavoidable, then building a better world involves understanding where stories fall short in their sense-making, e.g. in complex systems, and creating richer ones that inspire unified collective action [Bal15].

The potential applications of new insight into storytelling are far-ranging, from politics to healthcare to psychology. How stories influence people and impact society is clearly complicated to unravel. As David Kaiser said, "presumably more [is] involved than Austinian speech-acts."[5]

2 Methodology

2.1 Motivation

2.1.1 Why do we think it's reasonable to investigate these claims in space?

A classic approach to trying to understand a system is to consider how the parts of it relate to one another in space. As Waldo Tobler, paraphrased by Noel Cressie, said, "Everything is related to everything else, but near things are more related than distant things" [Cre21]. The spatial metaphor of learning about how things interact through their relative positions is a very simple, but hopefully powerful, tool to use to try to understand stories.

We think it is reasonable to use methods based on linear algebra to study human cognition, on both an individual and a collective scale, because (1) lots of other smart people already thought so [MR74; Rus80] and because (2) spatial metaphors are a well-attested feature of human cognition across topics [Cas17], cultures, and time periods [Hor+16]. Therefore supposing that people may use spatial metaphors in conceptualizing fictional characters is plausible.

[4] Sometimes we deal with utterances with respect to their truth value. In the realm of stories, this is obviously incomplete: we find meaning in stories without needing to evaluate their truth value in the logical sense. A brief aside on the magical or emotional value of utterances: language has multiple, competing uses that look superficially similar [Wt97], which range from illocutionary acts which directly change reality to poetry to offering pat phrases with touching sincerity. We think, besides the logical truth value, there's also an emotional and sometimes a magical value. For example, when someone is in a sinking boat and says, "this boat isn't sinking", they aren't necessarily lying - trying to persuade you of something they know is false - so much as performing an incantation: by speaking, they hope to change reality. It seems unlikely that we are perfectly able to distinguish between these gradations of speech, those that (mostly) change the world and those that (mostly) describe it.

[5] "Ideas alone do not force people to interpret, appropriate, or act in a coherent way. Ideas alone do not radically reshape lumbering institutions like the U.S. infrastructure for scientific research. We must interrogate how certain ideas (and not particularly new ideas, at that) came to seem convincing to particular people in specific times and places; how various people chose to act on them, tamping down variant interpretations, building coalitions, and overcoming inevitable turf battles and bureaucratic inertia. Presumably more was involved than Austinian speech-acts."[Kai13]

2.1.2 Why do we think it's reasonable to use dimensionality-reduction techniques like SVD and PCA?

There are a lot of statistics- and linear-algebra-related methods used to decide which "factors (the number of latent dimensions underlying the data generating process)" are important in a specific psychological context; answering which method is best is complicated. It seems like recent conception of best-practice is that using multiple methods and combining them via machine learning is often a solid approach, and can out-perform theoretically-based criteria. However, these models are difficult to interpret [GB20]. Since fundamentally what we're after is an interpretation of what it means to define a character, it seems appropriate to use dimensionality reduction techniques - namely singular value decomposition (SVD) and principal component analysis (PCA) - to come up with a model for the space, especially given prior research into human personality [MR74; BL94; AL01].

In future work, we'd like to further experiment with iteratively applying SVD and PCA. We think this approach may be worth exploring in general because of the method of Power Iterations (in SVD)[Bru21], and in particular based on the work of other researchers, who were able to use iterative processes to filter out the parts of a story they particularly wanted to study: "Therefore, we conduct regression for each feature to discover dominant tendencies of the features. By filtering scenes that do not follow the tendencies, we extract a story line that exhibits the most dominant personality changes. We can decompose stories into multiple story lines by iterating the regression and filtering."[L2K21]

Given the wide range of models (and models that are combinations of models) and approaches to pulling out the dimensionality of a space and the relative importance of the factors in that space, future work could involve showing whether or not applying more of these methods agrees with our preliminary results. For example, "[t]he degree of representativeness [of a prototype/ feature/ factor] can be measured using a distance function to a salient entity of the category..."[FP17], so although we initially used PCA and SVD to identify groups of characters and groups of traits (and their weights) to define the space, and interpreted from there, there are many other ways to approach interpreting the structure of the space. If our interpretation is on the right track, then if we applied these measures appropriately, and made sure their intentions aligned, then the results from any such process would be somewhat reconcilable with the results we obtained (and if not, we'd need to revisit our initial interpretation, since fundamentally all of these tools aim to unearth the same things).

One such avenue we would like to explore in future works is archetypal analysis, especially within individual fictional works or groups of works, which we could use to identify types of characters via an additional method that seems particularly well-suited to the study of characters. Archetypal analysis (AA) "...proposed by Cutler and Breiman [8], identifies the K groups with respect to a set of K extreme points, called archetypes, and aims to maximize the heterogeneity among the K groups... AA can be defined in terms of a factorization problem of the data matrix X under different constraints."[FP17]

2.1.3 SVD as a black box: 1000 foot view

If you, Dear Reader, are not very familiar with linear algebra, SVD, or PCA[6], we hope we can still describe our project sufficiently for it to be meaningful and interesting to you. If you'd like to skip the details of SVD and PCA, you can imagine their processes as a black box: we put in our initial data set, which is fictional characters and their average score for each personality trait differential, so each character is being described as a combination of around 250 aspects of personality, and what we get out is (hopefully) a way to more-or-less describe the same set of characters, but with far fewer personality traits. For example, if Elizabeth Bennet is judgmental, clever, witty, funny, caring, compassionate, acerbic, ... energetic, prideful, and sensitive, a list of 250 personality traits, then what we hope to get out from our analysis is a description of Elizabeth Bennet that is almost as good but much more

[6]Neither are we, really!

succinct: maybe something like, good, irreverent, loyal. The analysis won't come with those simplified labels (good, irreverent, loyal, in this example), but our goal is to infer fitting labels for the major axes based on the pattern of which traits are important to defining it, which is something the analysis will include.

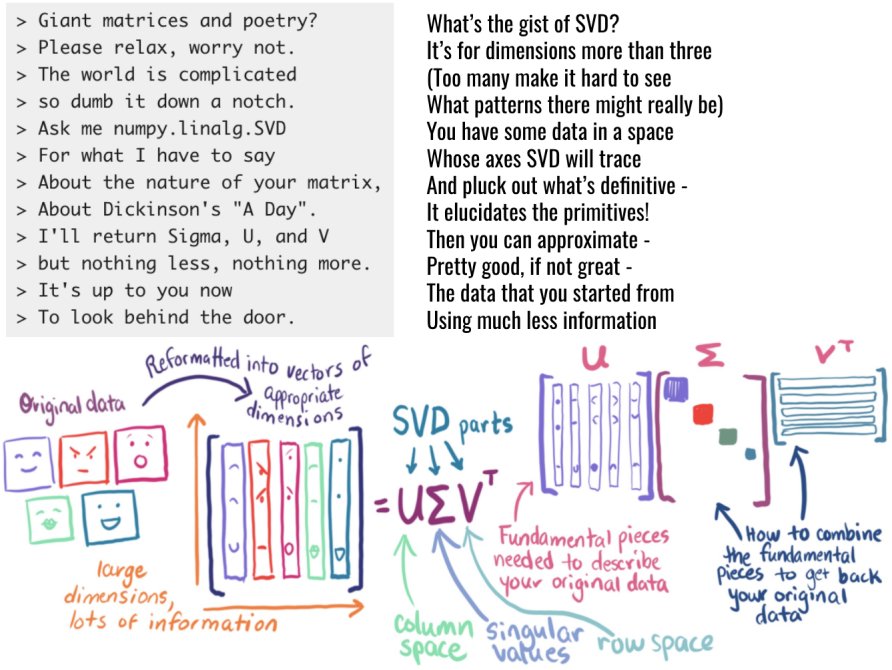

Figure 2: Some hopefully helpful explanatory aids.

2.2 Tools

We're using the Python programming language via Spyder IDE and Jupyter Notebooks; Slack messenger and various video chatting platforms for communicating with each other; Github for sharing and storing files[7]; Python packages including numpy, matplotlib, seaborn, pandas, scikitlearn, json, and others; and all sorts of resources from the internet including YouTube, Wikipedia, and Stackoverflow.

2.3 Source of our data

The dataset we're using comes from Openpsychometrics.org. Visitors to the site can take a number of psychology-related quizzes, and much of that data is then available for research. Our dataset comes from a quiz which pairs a user with their "most similar" fictional character. A user would assign a score for a personality trait differential to a given character by adjusting the slider on the screen. Each trait is described as a differential, so as a spectrum with one adjective on the left and a second antithetical adjective on the right. These anchor words, understood in opposition to each other, describe the trait being measured for that character. Most of the trait differentials were in a format similar to Figure X. We used all of the differentials that used words (we excluded emoji differentials). The scores are on a scale from 0 - 100, where 0 is the extreme left end of the slider, and 100 is the extreme right end of the slider. The slider is not very large on the screen by default, so the precision with which a user scores a character is not very high. As this dataset involves hundreds of characters and traits, the patterns that

[7]Some of our code is available at https://github.com/jwzimmer/tv-tropening

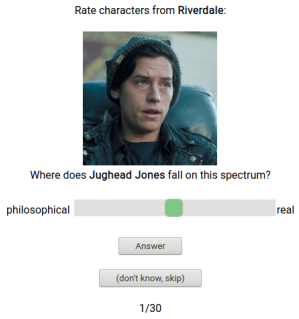

Figure 3: The example image given in the Open Psychometrics codebook, where users can move the slider to rate the pictured character with respect to a pair of traits.

might be within it are not easy to spot. We are using methods from linear algebra, e.g. SVD, in the hopes that we can distill some of the information in the dataset into a more interpretable format.

The main object we're manipulating with our code is a matrix of the average scores for each character for each personality trait. So along one axis are all of the fictional characters, and along the other axis are all of the pairs of personality traits, or trait differentials. The value of a single cell is the aggregated score of that character for that trait. By "aggregated score", we mean the average score across every person who took the quiz and provided a rating for that character and that trait. In other words, we used this data to create an 800 (characters) x 236 (trait differentials) matrix, where the average score for character i and differential j were the contents of the ij^{th} cell.

2.3.1 Possible limitations

There are a number of things about the underlying dataset which give pause to how generalizable any results based on it might be. For many included fictional works, such as Pride and Prejudice and Ender's Game, there are multiple versions (e.g. book and movies) that a participant might be thinking of, and the characters aren't guaranteed to be sufficiently similar from one presentation of the work to another. The character to be rated is presented with a picture on the site, which seems like it could influence participant's reactions. For example, Commander Graff from Ender's Game is portrayed in the movie by Harrison Ford, and his picture (in character) is used in the quiz. It doesn't seem far-fetched to think that a participant's judgments of other characters Harrison Ford is famous for, like Indiana Jones or Hans Solo, or whether or not a participant finds him attractive, could seep into their judgment of him as Commander Graff.

The possibilities for invalidity seem almost endless: participants are asked to self-identify which works they are familiar with, participants have to self-select to visit the site and take the quiz in the first place, and if participants have a particular beloved character in mind they want to be scored as similar to by the quiz, they might strategically answer questions with that end in mind. There's also a lot we don't know about how the data was gathered. For example, we don't know whether words in a differential were presented on the left or the right at random or always on the same side[8].

In short, we don't expect that the exact numerical score for each character is deeply meaningful, or likely to be consistent across participants. Philosophically, of course, not everyone is going to have the same interpretation of a character or a story, but we believe this doesn't mean there isn't enough universality in what stories mean to people for us to study them. There are statistical tests we could run on the data to try to nail down its properties, which may be part of future work, but for now we have chosen to work with it as is, and to keep its limitations in mind when interpreting it.

[8]We emailed the site email address to ask but have not yet heard back.

2.4 Data Cleaning

2.4.1 Choosing characters and traits by number of ratings

Because the number of ratings for each trait and each character varied, we considered taking out traits and characters that had too few ratings to be reliable. After removing some of the trait differentials as described above, our matrix was 800 characters x 238 traits. We looked at the distribution of number of ratings per character, per trait and saw that the distribution was right-skewed; it was more common for a trait, character pair to have a smaller number of ratings than a larger number of ratings. The average minimum number of ratings for a single trait per character was 42.77375 and the standard deviation for number of ratings was 52.89494. Using average divided by two for the minimum threshold, so about twenty ratings, that left 417 characters. While taking out characters with very few ratings might make our dataset slightly more consistent, since characters within the same work can have different numbers of ratings, when you look at an individual work you may be missing some important characters (for example, Wickham, from Pride and Prejudice). Therefore for most analysis we kept all the characters in, regardless of number of ratings.

We did not exclude traits or characters based on high standard deviations, because we thought it might exclude characters or traits which people find genuinely polarizing.

Promisingly, running sklearn's TruncatedSVD on the set of 417 characters versus 800 characters yielded similar output matrices, at least for the first few dimensions.

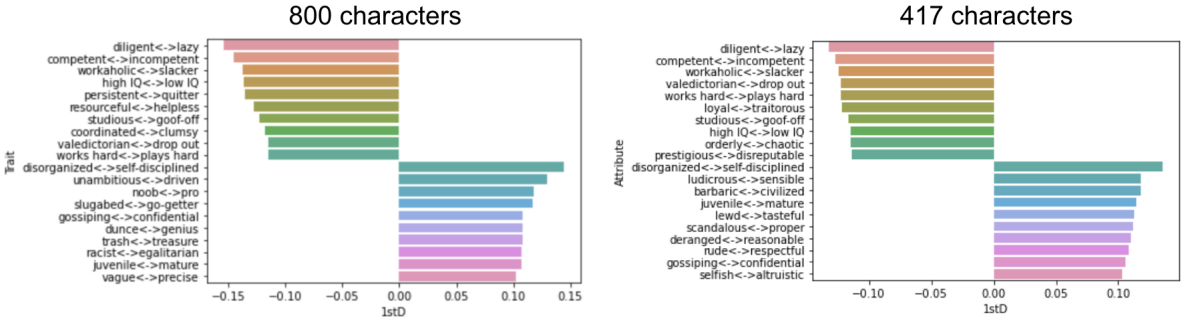

Figure 4: Viewing the traits with the largest magnitudes in the linear combination of the first dimension in the output matrix from TruncatedSVD.

2.4.2 Reorienting trait differentials

Although the output of the SVD algorithms we used, since they were matrices, were relatively interpretable, compared to many other options (like neural nets), they are still hard to read and make sense of (at least for us). Therefore we tried a few strategies that might make the output more easily human-readable. One such strategy was realigning the trait differentials to be more consistent; basically, to put the preferred trait consistently on one side, and the worse trait on the other. However, this turned out not to be straightforward to do! We ran into several issues, described in detail below, which we think all stem from one root issue: context.

(1) A consistent alignment may be impossible because there are words that appear multiple times as anchor words; because of context, a trait which is "preferred" compared to one trait may not be preferred in juxtaposition with another.

Using the Python "Counter" object from the "collections" module, we checked for repeated words in the pairs of trait differentials, and we also checked for loops (e.g. traits that share enough anchor words to form a loop, like "playful-shy-bold-serious-playful" from four differentials). We found that "playful"

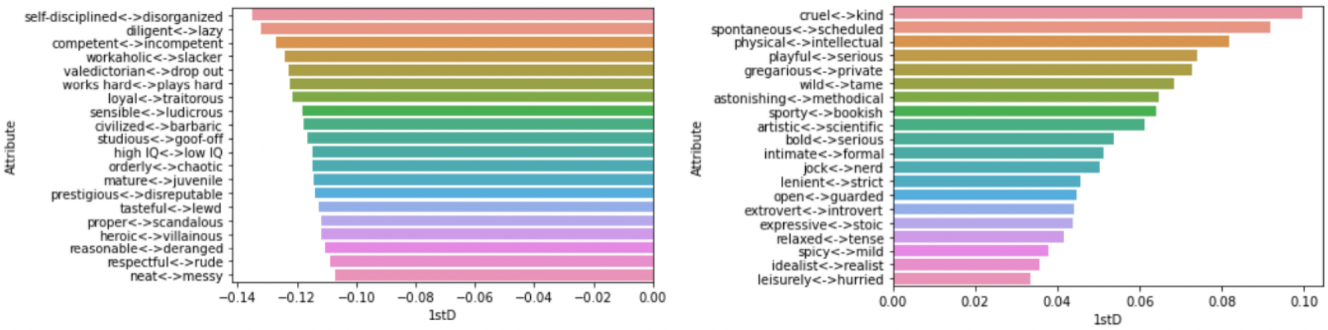

Figure 5: For comparison to the first dimension before realigning (in the previous figure), here are the traits with the largest magnitudes in the first dimension in the output matrix over 417 characters from TruncatedSVD with some trait differentials reoriented by hand (the word that was originally on the left now on the right, to make the output easier for us to read).

and "charming" occur as left anchors twice: in "playful<->serious", "playful<->shy" and "charming<->awkward", "charming<->trusting", and that "bold" and "suspicious" occur as right anchors twice: in "shy<->bold", "serious<->bold" and "trusting<->suspicious", "awkward<->suspicious".

The words which occur on both left and right as anchor words (if you consider direction) create loops between the differentials (if you don't consider direction, all the words which occur multiple times can create loops). Although we kept this in mind as a complication when trying methods for realigning the anchor words, we don't think this is enough basis to infer transitivity between any of the traits. The following trait differentials could be involved in loops:

```
['playful<->serious',
'shy<->bold',
'charming<->awkward',
'trusting<->suspicious',
'spontaneous<->deliberate',
'scheduled<->spontaneous',
'quarrelsome<->warm',
'playful<->shy',
'serious<->bold',
'charming<->trusting',
'awkward<->suspicious',
'warm<->cold']
```

(2) Due to societal norms[9] – for example, gender roles – a trait that is preferred in one character may be seen as a shortcoming in another. Although this makes aligning the trait differentials in a consistent way difficult, it points to a whole avenue of study. We plan to explore how sorting the characters by perceived binary gender impacts the kind of traits that are most important in describing the space, and in future work, we hope we might be able to get feedback from scholars in Women's and Gender Studies on this angle.

(3) Representations of words as vector embeddings don't sufficiently capture the difference in meaning between one of these anchor words alone versus when it appears as one end of a spectrum. As an alternative to aligning the traits by hand, we considered using a tool to map words to vectors of features in order to calculate a "distance" between them. Options included ConceptNet[SCH17], Glove, and word2vec. Obviously, NLP is imperfect, and there's no reason to expect all of the judgments derived from any set of NLP tools to perfectly agree with the judgments of a small-ish group of people; sometimes these differences are immaterial, but in this specific application that did not seem to be the

[9]We mean to be descriptive, not prescriptive, in mentioning stereotypes and heteronormative assumptions.

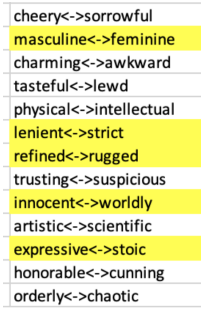

Figure 6: A few examples of traits whose desirability seems likely to be influenced by gender norms.

case.

When people are confronted with a pair of semantic differentials, part of what they get out of that context is that they should juxtapose the two words; they should try to interpret those two words as being as different from each other as possible. For example, "hot<->cold" should not be viewed as similar temperature words, but as opposite ends of a spectrum. Out-of-the-box NLP tools don't seem to be sensitive to this (unsurprisingly - not a criticism, just noting a potential limitation). For example, see the unexpectedly low score – meaning high similarity – of "masculine-feminine" below:

```
In [45]: distances

Out[45]: {'playful-serious': 0.588435709476471,
          'shy-bold': 0.6278842091560364,
          'cheery-sorrowful': 0.680909663438797,
          'masculine-feminine': 0.137395441532135,
          'charming-awkward': 0.5569109618663788,
          'lewd-tasteful': 0.4542396068572998,
          'intellectual-physical': 0.4425240159034729,
          'strict-lenient': 0.40918368101119995,
          'refined-rugged': 0.6156565546989441,
          'trusting-suspicious': 0.7093296349048615,
          'innocent-worldly': 0.7068087458610535,
          'artistic-scientific': 0.43224066495895386,
          'stoic-expressive': 0.6233098208904266,
          'cunning-honorable': 0.7130308151245117,
          'orderly-chaotic': 0.5125446021556854,
          'normal-weird': 0.47316068410873413,
          'competitive-cooperative': 0.477466344833374,
          'tense-relaxed': 0.5601309537887573,
          'brave-careful': 0.3970715403556824,
```

To get around this issue, we tried framing the alignment problem as an optimization problem, in which it does not (directly) matter how similar "masculine" and "feminine" are judged to be.

To expand on that, here is a pseudocode description of our planned algorithm. Say we only have 3 differentials in our dataset: good<->bad, hot<->cold, and frigid<-friendly. Visually, we can see that there is one differential we'd like to reorient. The one we'd like to change is frigid<->friendly to friendly<->frigid. Our algorithm should produce an outcome consistent with that intuition. We want to look at sum-of-distances-between-the-pairs(good, hot, frigid) versus sum-of-distances-between-the-pairs(good, hot, friendly). When switching a pair of words causes that score to go down, then we want to reorient that semantic differential. We chose to use pre-trained gensim/ Glove/ models.

(4) The spectrum spanned by a pair of anchor words could vary widely. Some pairs of traits, like "mathematical<->literary", are describing a spectrum that does not encompass a wide range of behaviour. By contrast, the difference between "cruel<->kind" could be vast, considering the range of historical atrocities and imaginary rulers in fictional works. Trying to form a consistent alignment implicitly conjures a fungibility across the trait differentials that can be misleading.

Ultimately, using the word embeddings did not produce an alignment consistent with our intuition, and once we explored realigning the words by hand, and realized it was likely there was no consistent way to align the words the way we initially imagined, we used the original alignment. Given that realigning the anchor words is a relatively small sub-task within our larger project, we wanted to balance these concerns with the appropriate scope.

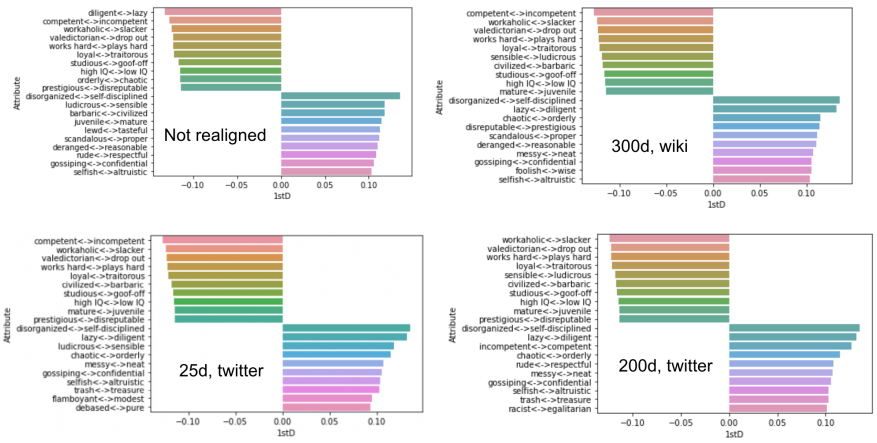

Figure 7: Output after using word embedding models to realign the trait differentials (first dimension shown).

2.4.3 Normalization

Initially, we left the data raw for PCA and renormalized the data for TruncatedSVD so that it would be centered at 0 and between -1 and 1 (instead of 0 and 100). However, we decided that might artificially compress traits in a way that was misleading, making traits appear more similar than they actually are, so by the time we implemented SVD using linalg.svd, we were leaving the data untouched.

In future work, we think it would make sense to continue exploring removing the mean per trait and, in some contexts, normalizing for spread in the first few dimensions per work in order to explore the analogy with vowel space. The reason we started considering removing the mean is because that might take some of the burden of explaining a facet of personality that all characters have in common out of the SVD-based approximation, allowing the differences between characters to make up relatively more of the information captured by the decomposition.

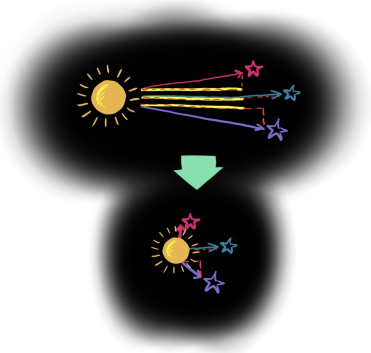

Figure 8: A visual analogy for why we want to remove the mean for each trait during SVD: if you consider how far three stars are from the sun, a big chunk of the information captured by drawing vectors to them is the horizontal distance they have in common. When this distance is included, the three vectors look relatively similar to each other. When this distance is removed, only the portion of the vectors that are different from each other remain.

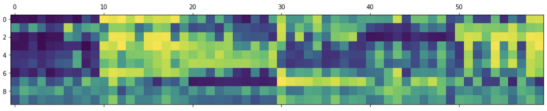

Figure 9: The scores (normalized between -1 and 1 and centered at 0) for the (approximately) top-10 most-positively-weighted and top-10 most-negatively weighted traits from the first 6 dimensions of the output matrix from TruncatedSVD, for the first ten characters ("zoomed in" for visibility).

2.5 Singular Value Decomposition (SVD) and Principal Component Analysis (PCA)

Note that the parts of the matrix decomposition produced by SVD are almost unique, except that their signs (as in, positive or negative) could flip without impacting the reconstruction of the original matrix. So we should be cautious in interpreting the meaning of the signs in the output of these processes.

2.5.1 SVD

SVD is a linear dimensionality reduction method that attempt to find linear combinations of features of a data matrix so that the information represented in the original data can be represented nearly as well with only a subset of the data.

Suppose we have a matrix X of size $M \times N$. SVD decomposes this matrix into the product of three factors where

$$X = U\Sigma V^T.$$

U and V are orthogonal matrices. The columns of U contain the left singular vectors of X, the rows of V^T contain the right singular values of X, and along the diagonal of Σ are the singular values of X ordered by descending importance.

What's important is that we can use U, Σ, and V^T to reconstruct X as the sum of rank 1 matrices such that

$$X = \sigma_1 u_1 {v_1}^T + \sigma_2 u_2 {v_2}^T + \cdots + \sigma_p u_p {v_p}^T$$

where $\sigma_1 u_1 {v_1}^T$ explains the most variance, followed by $\sigma_2 u_2 {v_2}^T$, and so on up to $p \leq N$. Because the trailing terms explain the least variance, they can be truncated without a significant loss of information. The result is a lower-rank matrix, say a rank 3 matrix $\tilde{X}$, that well approximates X:

$$\tilde{X} = \sigma_1 u_1 {v_1}^T + \sigma_2 u_2 {v_2}^T + \sigma_3 u_3 {v_3}^T$$

2.5.2 PCA

PCA is similar to SVD. The goal of PCA is to find linearly uncorrelated orthogonal axes, i.e. the principal components, of a covariance matrix onto which we can project our data points. Performing eigendecomposition on a covariance matrix C returns

$$C = W\Lambda W^{-1}$$

where W is a matrix containing the eigenvectors of C and their corresponding eigenvalues contained in the matrix Λ. The principal components are the eigenvectors arranged by the magnitude of their eigenvalues such that the first principal component explains the most variance, followed by the second principal component, and so on. These principal components provide a basis for a lower dimensional coordinate system onto which a data matrix can be projected without a significant loss of variance.

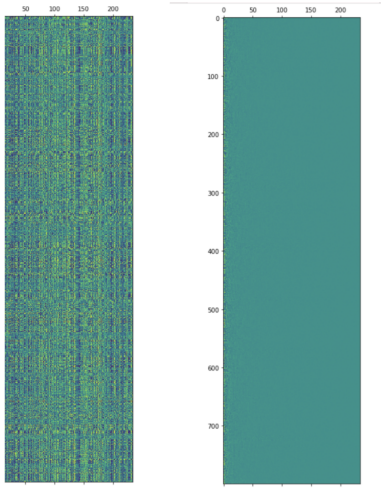

Figure 10: Before (left) and after (right) using sklearn's TruncatedSVD on our characters x traits matrix for all 800 characters. 15 dimensions explain about 87% of the variance.

So if we have a data matrix X, then a k-th dimensional representation of X where k is less than the original number of features can be found as

$$X_k = XW_k$$

where W_k is a matrix with k principal components.

3 Results

There are different ways to try to reduce the dimensionality of a space, and even within the SVD/ PCA realm we chose, there are multiple implementations of the algorithms available. However, if the implementations are reasonable, and our application of them is reasonable, then despite some differences, they should lead to approximately the same results. Since one of our main goals in this paper is establishing whether or not this kind of spatial analysis of character composition is meritorious enough to warrant future work, we tried related variations on the same theme; if our assumptions are totally off base, then we might see unexpected differences in processes we expect to produce similar results.

In our case, for SVD we chose to use scikitlearn's TruncatedSVD, which returns a reduced-dimension approximation of the original data matrix, and numpy's linalg.svd, which returns (basically) U, Σ, and V from the definition of SVD (see the Singular Value Decomposition section for more details). The output returned by TruncatedSVD is similar to a combination of the three kinds of output returned by linalg.svd, so when we talk about a dimension from TruncatedSVD, that includes information from U, Σ, and V all put together, whereas when we talk about about an element from linalg.svd, that will be a vector from U, a singular value from Σ, or a vector from V, a single part of the original data's decomposition.

3.1 Linear Regression model

We used sklearn's LinearRegression model to see if the "orderly<->chaotic" trait differential score for a fictional character could be reasonably predicted based on a subset of the most-extremely-weighted traits in our first dimension (the dimension that explains the most variance after passing the character x trait matrix through TruncatedSVD), because if the SVD process is meaningfully drawing out the

structure of the space of character traits, then traits within a large dimension should be fairly predictive of each other[Cre21]. If the prediction fit the ratings exactly this would be the line y=x. We didn't

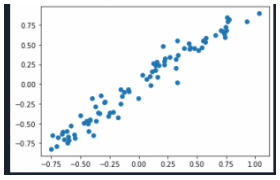

Figure 11: The predicted scores vs. the actual scores on the validation set. The model had a Training Accuracy of 0.961 and a Validation accuracy of 0.955.

expect a perfect fit, but this looks close enough to a line to think the relationships between the traits we proposed are not implausible. A linear regression model should not perform as well when it predicts traits across dimensions, or when it uses low-weighted traits to make predictions; this may be something to verify in future work.

3.2 Overall

We used PCA to find clusters of the characters by their approximate reduced-dimension matrix approximation. Note that PCA and SVD are similar processes based on the same concepts from linear algebra, and that depending on the details of the implementation of PCA or SVD you're using, you may be able to convert between one matrix approximation and the other.

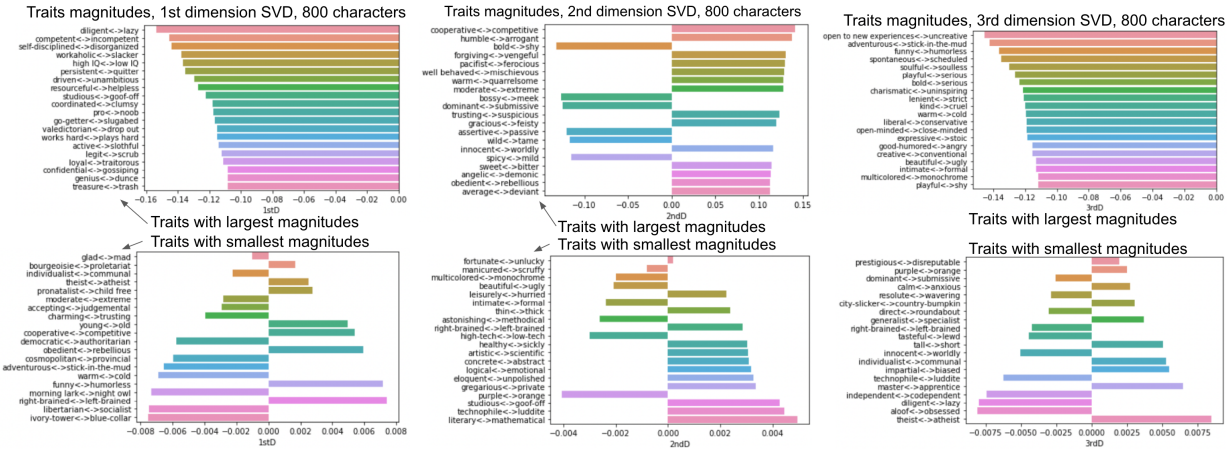

Figure 12: The traits with the smallest and largest magnitudes in the first three dimensions from the output of TruncatedSVD (with hand-realigned trait differentials).

3.2.1 Initial Description

We have barely scratched the surface of what this space looks like, but it does look like some patterns are emerging from it, so we plan to continue this research. If we look at the characters clustering at the high end of the first principal component in Figure 12, some of those characters – like Baron Harkonnen and Joffrey Baratheon – scare us because their viciousness is stronger than their substantial appetite for power, which makes them both unpredictable and dangerous, whereas some of the characters in the bottom half – like Jane Bennett and Alfred Pennyworth – are humbly selfless, reliable, and predictable, which is comforting. For example, when we look at a specific work such as Avatar: The Last Airbender,

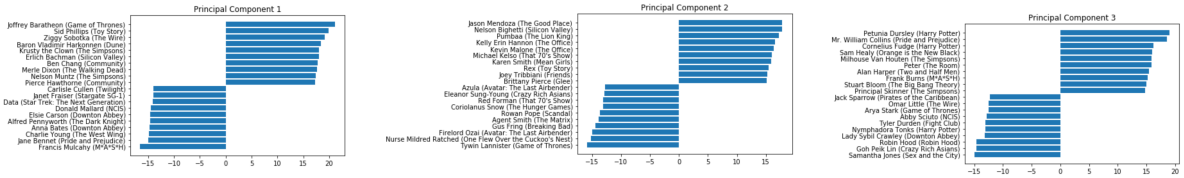

Figure 13: The characters with the highest trait scores in the top 3 principal components across all works.

we see this trend emerge (Figure 13). Azula commits a significant act of betrayal against her brother, Zuko, in one of the defining episodes. She is the daughter of Firelord Ozai, the main villain of the series, who is dangerous and evil, but more predictably so than Azula. Katara, on the other hand, is a consistent source of refuge and goodness, in stark contrast to Azula. Maybe, then, the most significant axis in character space is something like safety<->danger.

If we look at the second principal component across all works, several of the top characters provide comic relef – like Pumbaa, Karen Smith, and Joey Tribbiani – and we do not expect them to be particularly effective in carrying out their own plans; that isn't the role they have been constructed to play within the story. Characters in the bottom half are potential threats because they are powerful, authoritarian, and deliberate – like Coriolanus Snow, Tywin Lannister, and Gus Fring. In Avatar: The Last Airbender, Aang is a comedic, fun and lighthearted character, while Firelord Ozai and Azula assert a sense of dominance and authority (however Aang is certainly powerful in his own right – he is after all the Avatar). So maybe the second most significant axis in the space has to do with power and control, both over others and yourself.

Looking at the third principal component, the top half includes characters we find contemptible because of their purposeless, morally-bereft rigidity – like Petunia Dursley and William Collins (cogs in the machine) – whereas the bottom half includes characters who openly buck society's expectations – like Robinhood and Arya Stark (rage against the machine). This description fits the oscillating nature of characters in the bottom such as General Iroh, Ty Lee, and Aang. On the flip side, Firelord Ozai is singularly devoted to defeating Aang and controlling the world; there isn't any sense of chaos in his pursuit. Perhaps this third axis is something like lawful<->chaotic. An important step in continuing

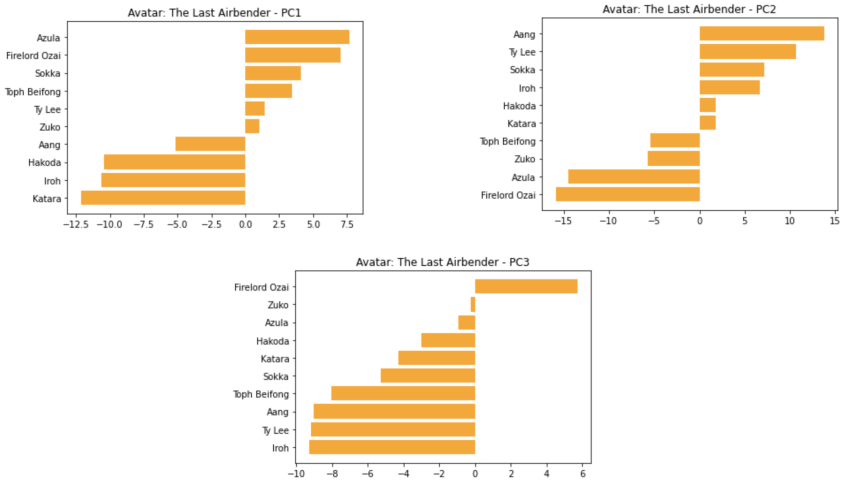

Figure 14: The characters with the highest trait scores in the top 3 principal components within the work "Avatar: The Last Airbender".

this research will be familiarizing ourselves to some extent with more of the characters in the dataset, so that we can verify is these patterns generalize or if they are too circumscribed based on the characters we already happen to be familiar with. We would like to continue comparing our results to descriptions

Figure 15: One way to describe the first three dimensions from TruncatedSVD might be Peril, Power, and Play.

of archetypes from literary criticism and other fields. With respect to the characters clustered at the bottom of the first principal component, we found a strikingly apt description from Tolkien about the purest kind of heroism: "the truly heroic situation, says Tolkien, was that of Beorhtnoths soldiers. "In their situation heroism was superb. Their duty was unimpaired by the error of their master." Consequently, "it is the heroism of obedience and love not of pride or willfulness that is the most heroic and the most moving.""[Gra73] We think our application could be used to support or refute extant proposals about literary archetypes which previously have been solely qualitatively observed.

3.2.2 Output from SVD processes

There are multiple options in Python for implementing SVD using pre-existing tools (as opposed to from scratch). We initially used **sklearn's TruncatedSVD**, which does involve stochasticity as its an implementation of randomized SVD intended to allow efficient completion of this process on large datasets. Because that's its intended use case[10], the output it returns is a truncated approximation of the original data (the economy SVD), rather than the individual components from the SVD formula. This makes it very easy to work with, but not quite as transparent as we might like. Because our dataset is quite small, completing every step of the SVD process is not prohibitively slow, so we also used **numpy's linalg.svd**, which returns U, a list of singular values, and Vh, when we recognized that we wanted to be able to pull apart the pieces of the decomposition more than we were able to using the economy SVD. A convenient feature of **numpy's linalg.svd** is that we can verify that it's working by reconstructing the original data matrix from U, Σ, and V. If we apportion a total of 1 to the singular values (normalizing them), then the first singular value makes up about a quarter of that total, 0.256, followed by a precipitous drop as the next largest singular value makes up about 5%, 0.0534. Singular values 3 through 10 steadily decline: 0.0465, 0.0368, 0.0272, 0.0218, 0.0166, 0.0162, 0.0153, 0.0120 respectively. Note that the difference in importance from one singular value to the next continues to get smaller; the difference in importance between the first few dimensions is much bigger than the difference in importance between the latter dimensions. This makes sense, because SVD decomposes the original data by its most significant features. We can think of an analogy with money: there is a difference in the cost of going out to eat or going to a movie with friends (the last few dimensions), but when you are trying to account for your expenditures, those will pale in comparison to your college tuition payment (the first dimension). The size of the first singular value (about 21890) tells us that

[10]Tentatively, since the stochastic version produced consistent results, that could be a good sign that the method of SVD was appropriate to apply in any guise in the first place?

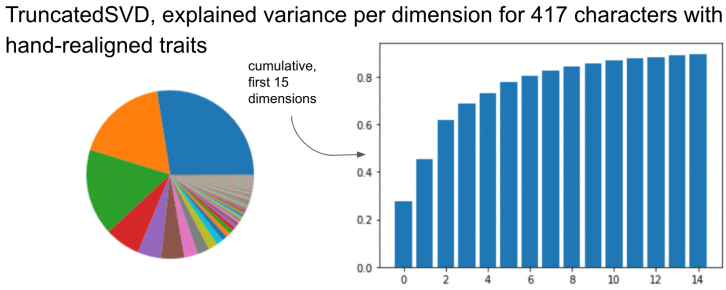

Figure 16: Looking at how much variance is explained by each dimension of the TruncatedSVD approximation of our original data, we can see that they get smaller fairly quickly. The first 3 dimensions get us over 50% of the way towards describing our original data.

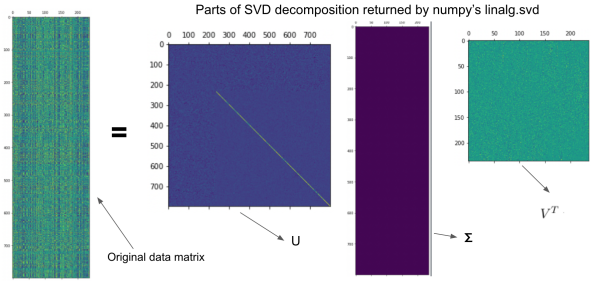

Figure 17: The output returned by numpy's linalg.svd which can be recombined to make the original data matrix.

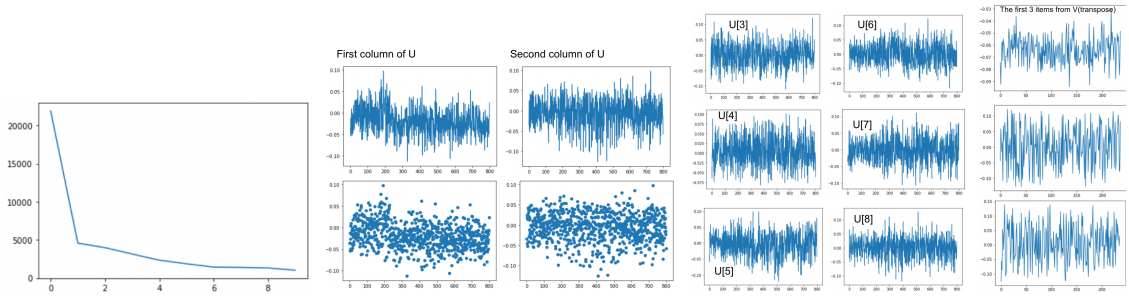

Figure 18: At left, looking at the size of the singular values returned by numpy's linalg.svd decomposition, we can see that the first one is by far the largest, and that they get small fairly quickly. In the middle, we can see the first few columns of the matrix U returned by linalg.svd, and at the right, the first few items from V^T.

the first part of our SVD approximation for our original data is almost five times as important as the second part (whose corresponding singular value is about 4570) for describing our original data.

3.3 Pride and Prejudice

We can look at one work at a time to get a sense for whether this interpretation of the character space makes sense through a specific familiar lens. We used Jane Austen's Pride and Prejudice because it's a popular, well-known work, both in novel form and as several movie adaptations, and because of that, it has been used in related research, e.g. Hedonometric analysis[DDC21],[Rea+16],[PRL20], and we looked at in the context of the TV Tropes website in a previous project. The character ratings in our dataset

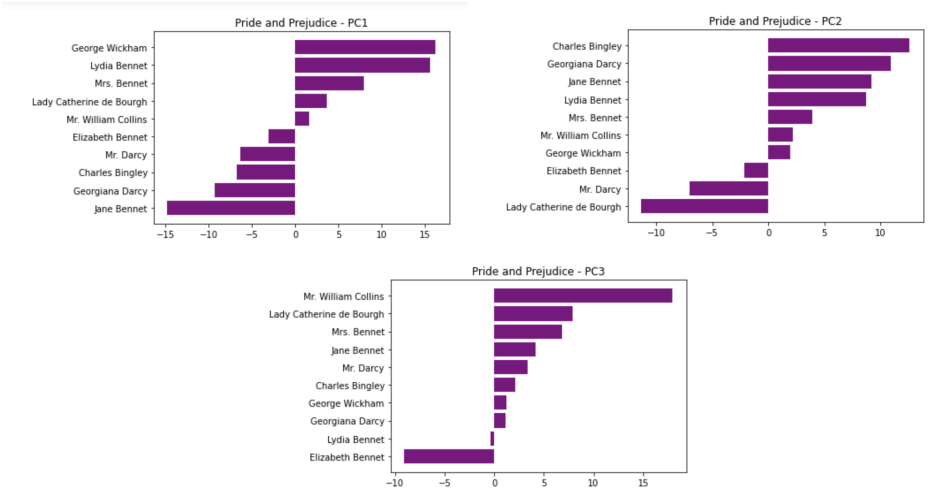

Figure 19: The characters with the highest trait scores in the top 3 principal components within the work "Pride and Prejudice".

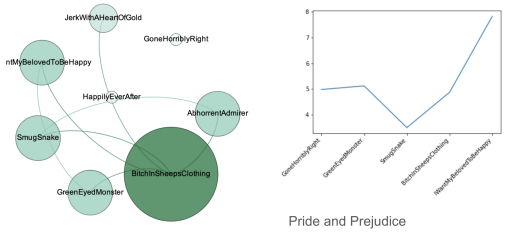

Figure 20: In a previous project, we looked at which tropes from the website TV Tropes appear in the synopsis of Pride and Prejudice. We used this information in conjunction with word happiness scores[DDC21] to get a rough picture of the story arc: things start off fine, there's a problem in the middle, and then a happy ending.

don't account for changes in the characters' personalities over time, so that could be an interesting facet to try to unpack in future work. For example, in Pride and Prejudice, the characters Wickham and Darcy undergo a significant change from Elizabeth's perspective at around the same time: she learns things to the detriment of Wickham and the benefit of Darcy, so we might expect to see some of their traits crossover each other at that point in the story. For now, we're limited to scores that reflect participants' judgments of the character overall, but there's still plenty to see: of note, if we arrange the characters in space by their scores from the first three dimensions from TruncatedSVD, there's a perspective from which some of the characters who end up together in romantic partnerships are paired together in the space – Jane Bennett and Charles Bingley are near each other, and Mr. Darcy and Elizabeth Bennett are near each other. The other pair in that graph, Lydia and Mrs. Bennett, don't end up in a romantic partnership (they're mother and daughter), but they do get paired together in their behaviour throughout the book, and particularly at the end; Lydia elopes foolishly with Wickham, and Mrs. Bennett is foolishly happy when they end up married.

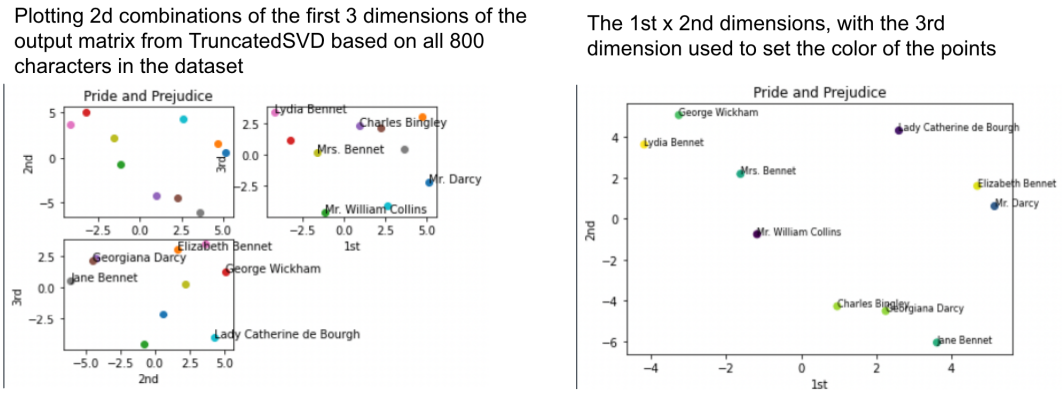

Figure 21: Using sklearn's TruncatedSVD output to visualize Pride and Prejudice.

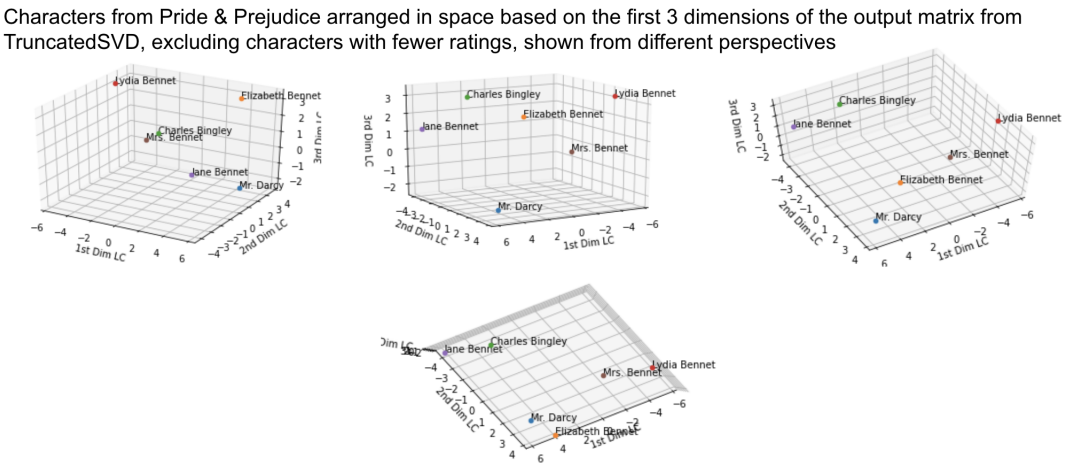

Figure 22: Using sklearn's TruncatedSVD output to arrange characters from Pride and Prejudice in space based on their scores in the first three dimensions.

4 Future Work

4.1 Proximal future work

The first priority of our plans for continued study is to thoroughly describe the features of the underlying space. Based on our initial exploration, we believe this way of looking at how characters are constructed will be a fruitful one. Describing the space in detail can be broken down into five categories: exploiting our initial methods, visualizing the space, finding the most helpful perspective from which to view the data (normalizing and rotating), applying other methods for revealing structure, and finding illuminating analogies. This last category is particularly open-ended, and we would like to try out ideas from other fields, but at least initially we think an analogy to linguistic vowel space could be productive.

4.1.1 Exploiting SVD, PCA

Although we have gotten a glimpse at what character space might look like, we have barely exploited the explanatory potential of SVD and PCA[11]. For example, we have not fully investigated the role of

[11]Professor Steve Brunton at the University of Washington has made some extremely helpful resources available for this at his book's website.

(the output matrices) U and V as "eigentraits" or "eigencharacters", determined if some portion of the matrix approximation reflects what characters have in common versus what distinguishes them (this also has to do with normalization and rotation), or compared in a systematic way how different levels of truncation (of the matrix approximation) affect the results. In particular, we plan to use power iterations[Bru21] and optimal truncation[GD14] to try to distill the most relevant character traits for each axis[12], by looking at what effect these operations have on the corresponding correlation matrices (and various measures of correlation, like Pearson R correlation or Spearman Rank correlation).

We plan on trying an idea suggested by UVM's Colin Van Oort to interpret the dimensions extracted by PCA/SVD. Since each dimension has the built-in interpretation of selection between two traits, we could, for each dimension that we want to describe, grab the two words associated with each dimension and pass them through a word embedding model to get word vectors; interpolate between the two word vectors using the PCA weight; sum all of the interpolated word vectors for the n^{th} PCA component, and then find the k concrete words that are closest to that point in word-embedding space. This should produce a handful of words that summarize a particular PCA/SVD dimension while accounting for the relationship between the words, and the weight observed in the PCA/SVD.[Oor21]

Part and parcel of exploiting the methods we've already started using is understanding the statistical properties of our dataset. We need to look at the distributions of scores and whether they are Gaussian (we have assumed they are close enough for now in applying these methods), and because it seems plausible to think that some of the responses in our initial dataset could be nonsensical – unreliable because e.g. a participant wanted to match a certain character or because they wanted to get through the quiz as fast as possible – we might need to apply a method like Robust PCA to try to de-bias our data[Can+09]. Finally, we need to see what we can learn from more of the dimensions than just

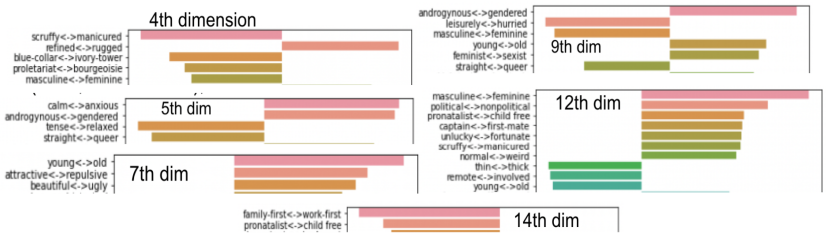

Figure 23: Traits that look related to gender roles and sexuality appear as some of the traits with the largest magnitudes in multiple dimensions of the output matrix from TruncatedSVD.

the first few. We have seen some hints of a gender/ sexuality/ procreative theme in some of the later dimensions from TruncatedSVD, so we need to unravel that. Since we have all the character names, we can divide them by perceived binary gender and re-run SVD on each group separately.

4.1.2 Visualizations

As mentioned above, we need more visualizations of the output of the SVD and PCA processes, including visualizations of more dimensions, visualizations of the basis functions, visualizations of which characters and which traits go together (for example, it would be helpful to see which are the most defining traits for the characters in a few example works, such as Pride and Prejudice), visualizations of individual works and authors' aggregated works, and which traits are significant across multiple dimensions.[13] We

[12] We made up a method for doing this, which we have been referring to as iterative SVD, because we did not know such a method already existed; it may be interesting to explore whether our method was doing what we intended it to, but that is distal future work.

[13] There are also a number of tools and chart styles which UVM's Computational Story Lab has developed or used which would be helpful, such as Professor Peter Dodd's ousiometer and word-shift charts showing changes in rank.

also need to create visualizations for which traits remain and which get "boiled off" by processes like power iteration.

4.1.3 Normalization and rotation

In addition to what was described above, we'd like to consider the interpretation of the orientation of the space: the linear algebra algorithms are picking up on which characters seem most different from our perspective, and the orientation of the information in the original matrix can make a big difference on the rank of the approximation needed to provide a faithful representation[Bru21], and we're not sure how those two factors interplay. The points -1,-1 and 1,1 are most distant from each other within a square centered at 0,0, but along an isolated axis, 0,-1 and 0,1 are the most distant, so the number of dimensions being considered changes how we should interpret distance.

4.1.4 Clustering, other kinds of matrix decomposition

There are many metrics and algorithms available for analyzing whether a dataset exhibits structure. We chose PCA and SVD initially, but there is no reason to think other methods wouldn't be illuminating as well. In future work we would like to see which characters are identified as being similar by clustering algorithms, to see if it is consistent with our initial results. We would also like to experiment with normalizing an individual work of fiction within a space as part of fleshing out how individual authors use the space and the analogy with vowel space. We would like to use other kinds of matrix decomposition as well, e.g. QR decomposition or Cholesky decomposition, to analyze the space.

4.1.5 Analogy to vowel space

We think characters, at least within stories, are constructed to occupy interesting, distinct parts of character space, akin to the way vowels are constructed within languages to maximize contrast.

4.2 Distal future work

There are innumerable interesting threads to follow with this project! For instance, we'd like to compare character types and character space to psychological concepts, to see whether these (e.g. the Dark Triad) and "common" personality types found in other research are reflected in fictional characters [GM17]. Even if some of the psychological concepts are considered outdated, it would be interesting to see to what extent they're embedded in literature, and to try to trace back as much as possible the chicken/ egg ordering. For example, can we make predictions of character outcomes based on where they are situated in character space, and the context of the story? Incorporating personality characteristics with metadata and context may improve prediction accuracy [AMS18].

It would be interesting to explore in more detail how individual authors utilize the space. If we have a theory for how a given author is using characters within a work to support their story telling, then given any $n-1$ of the main characters, we may be able to predict the approximate location of the n^{th}. We could investigate how morality tales like Candide may portray characters with extreme traits versus how gritty noir tales may cluster characters closer together, and depict no "purely" good or bad characters, or we could look at how Jane Austen constructs her novels around pairs of characters and character traits.

This dataset could be compared to many others[14]. We could see what blockbusters and bestsellers look like in terms of character space, or what sort of character portrayals get good audience or critical reactions. We could see if another set of personality traits, fictional characters, or ratings produces

[14]UVM's Professor Peter Dodds is working on a project involving word meaning which we plan to compare these results to.

similar results; this would be a way to support the interpretation of character space along a few main axes.

We'd also like to connect characters and character space with emotions. As Kurt Vonnegut said in 1981, "stories are made interesting through emotions that connect the characters, their motivations, goals, and achievements." This is a fertile area for exploring connections with other fields: cognitive scientists have pinpointed the central role of emotions in storytelling (Parkinson and Manstead, 1993; Hogan, 2011)" [BC20], psychologists and social scientists have broken down meaning in emotional terms like valence [Rea+16], and historians and classicists have studied how emotion is embodied in different schools of literature. In particular, we would like to pursue feedback from UVM's Romance Languages' Charles-Louis Morand Metivier, who studies emotions in French Medieval literature, because a comparison of how historical and cultural influences change how we devise characters seems like an exciting and rich topic.

Additionally, we'd like to compare characters in fan-fiction to mainstream fiction, since previous research has found surprising differences[PRL20], and because fan fiction is a bridge between the work itself and the broader response to it. Similarly, we'd like to examine readers' reactions to different character types: "By examining reader response at scale we can set the ground for statistically valid claims about how a lot of people read. Namely, we can understand what kind of emotions readers feel towards the characters and how they react to specific narrative strategies like suspense."[PRL20].

Finally, we'd like to connect character composition and archetypes to our exploratory research into tropes (based on the TV Tropes website), to see for example what kind of character traits are invoked with which tropes.

Figure 24: What kind of traits are associated with the "damsel in distress" trope?

5 Discussion

We should note that this project only touches on one aspect of character – personality – but of course there are other important facets of character composition like their body of actions and their social network, and of course characters can change and develop over time. Our peek into character space is necessarily not a complete representation of how characters work or what they mean.

Bridging the world of fiction and non-fiction, and bringing a somber bent to a fun topic, the impact of stories can persist despite evidence in the world indicating that the story is unlikely to be true. As we've seen time and again, disconfirmation doesn't necessarily mean much to continued belief[FRS56]. It might be that the way to fight society's ills is not through providing evidence of claims, but by

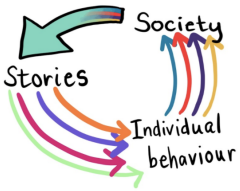

Figure 25: Stories consolidate individual behaviour and beliefs to an easily transmissible format on a societal level, which can be applied to each individual equally easily. A "many-to-one" operation followed by a "one-to-many" operation is more efficient than a "many-to-many"/ many "one-to-one" operations, like homogenization of coordinates in linear algebra.

providing the most compelling story[15]. Therefore studying how people conceive of and react to other people, even in a fictional context, has genuine real-world applications, both because fictional characters have recognizably human personality traits, and because we inevitably use stories to portray real-world events.

6 Acknowledgements

Thank you to UVM, Mass Mutual, and Professors Dodds and Danforth for the opportunity to be in school again. Thank you to Dave Dewhurst, Professor Guillermo Rodriguez, Josh Minot, Kelly Gothard, Colin Van Oort, Jane Adams, Nicholas Cheney, Ari Kotler, and everyone in the Computational Story Lab for their helpful feedback.

We have the resources and opportunity to work on this project due to luck and privileges granted by society, some of which come at the cost of other people and the environment. Therefore some remuneration is within the scope of this project, so we donated fifty dollars to the Environmental Defense Fund and fifty dollars to the American Civil Liberties Union.

[15]For example, regarding QAnon, "...our named entity detection analysis shows that Q drops refer to political matters as we detect words like the US, Hussein, the Senate, and House mentioned frequently. Also, it is evident that Q drops focus on agencies and institutions with almost 50% of the Q drops mentioning at least one. This is not surprising as the conspiracy theory discusses how agencies are controlled by the so-called deep-state and infiltrate the government to affect policy."[Ali+21]

References

[AL01] Michael C. Ashton and Kibeom Lee. "A Theoretical Basis for the Major Dimensions of Personality". In: *European Journal of Personality* 15 (2001), pages 327–353.

[Ali+21] Max Aliapoulios, Antonis Papasavva, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. "The Gospel According to Q: Understanding the QAnon Conspiracy from the Perspective of Canonical Information". In: *arXiv:2101.08750 [cs.CY]* (2021). https://arxiv.org/abs/2101.08750.

[AMS18] Danny Azucar, Davide Marengo, and Michele Settanni. "Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis". In: *Personality and Individual Differences* 124 (2018), pages 150–159. ISSN: 0191-8869. https://doi.org/https://doi.org/10.1016/j.paid.2017.12.018.

[Bal15] Philip Ball. "The Story Trap". In: (2015). https://aeon.co/essays/why-story-is-used-to-explain-symphonies-and-sport-matches-alike.

[BC20] Faeze Brahman and Snigdha Chaturvedi. "Modeling Protagonist Emotions for Emotion-Aware Storytelling". In: *Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)* (2020). https://www.aclweb.org/anthology/2020.emnlp-main.426.pdf.

[Bee98] Gillian Beer. *Introduction. Persuasion. By Austen, Jane.* 1998, pages x–xii. ISBN: 978-0-140-43467-5. https://archive.org/details/persuasion00aust_1/page/n21/mode/2up.

[BL94] Margaret M. Bradley and Peter J. Lang. "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential". In: *Journal of Theoretical and Experimental Psychiatry* 25 (1994), pages 49–59.

[Blo+20] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (Technology) is Power: A Critical Survey of "Bias" in NLP". In: *arXiv:2005.14050v2 [cs.CL]* (2020). https://arxiv.org/abs/2005.14050.

[Bru21] Steve Brunton. *Singular Value Decomposition YouTube Playlist.* A series of video lectures by Steve Brunton based on his book, available at http://databookuw.com/databook.pdf. 2021. https://www.youtube.com/playlist?list=PLMrJAkhIeNNSVjnsviglFoY2nXildDCcv (visited on 02/05/2021).

[But90] Jerome Butler. *Acts of Meaning: Four Lectures on Mind and Culture.* Harvard University Press, 1990. ISBN: 9780674003613.

[Cam08] Joseph Campbell. *The Hero with a Thousand Faces.* New World Library, 2008. ISBN: 9781577315933.

[Can+09] Emmanuel J. Candes, Xiaodong Li, Yi Ma, and John Wright. "Robust Principal Component Analysis?" In: *arXiv* (2009). [cs.IT]. https://arxiv.org/abs/0912.3599.

[Cas17] D. Casasanto. *Metaphor: Embodied cognition and discourse: The hierarchical structure of mental metaphors.* In B. Hampe (Ed.) Cambridge: Cambridge University Press, 2017, pages 46–61. https://casasanto.com/papers/Casasanto_2017_HMMT.pdf.

[CG16] Brian Christian and Tom Griffiths. *Algorithms to Live by: The Computer Science of Human Decisions.* 2016. https://algorithmstoliveby.com/.

[CM91] Joseph Campbell and Bill Moyer. *The Power of Myth.* Knopf Doubleday Publishing Group, 1991. ISBN: 9780307794727.

[Cre21] Noel Cressie. "A few statistical principles for data science". In: *arXiv* (2021). https://arxiv.org/pdf/2102.01892.pdf (visited on 02/04/2021).

[DDC21] Peter Dodds, Chris Danforth, and their team in the Computational Story Lab. *Hedonometer.org*. More information on how this tool was constructed at its website and in the paper Temporal Patterns of Happiness and Information in a Global-Scale Social Network: Hedonometrics and Twitter. PLoS ONE, 6, e26752, 2011. 2021. https://hedonometer.org (visited on 05/03/2021).

[Erw17] Martin Erwig. *Once Upon an Algorithm: How Stories Explain Computing*. 2017. ISBN: 9780262036634. https://mitpress.mit.edu/books/once-upon-algorithm.

[FC96] Jill Freedman and Gene Combs. "Narrative Therapy: The Social Construction of Preferred Realities". In: *Psyccritiques* 42 (Jan. 1996).

[FP17] Mario Fordellone and Francesco Palumbo. *Finding Prototypes Through a Two-Step Fuzzy Approach*. Chapter in Data Science. Springer, 2017, pages 111–121. https://dokumen.pub/data-science-innovative-developments-in-data-analysis-and-clustering-1st-ed-331955722x-978-3-319-55722-9-978-3-319-55723-6-3319557238.html.

[FRS56] Leon Festinger, Henry W Riecken, and Stanley Schachter. *When Prophecy Fails*. University of Minnesota Press, 1956. ISBN: 978-1-617-20280-3. https://doi.org/10.1037/10030-000.

[GB20] D. Goretzko and M. Bühner. "One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis". In: *Psychological Methods* (2020). Advance online publication. https://doi.org/10.1037/met0000262. https://pubmed.ncbi.nlm.nih.gov/32134315/.

[GD14] M. Gavish and D. L. Donoho. "The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$". In: *IEEE Transactions on Information Theory* 60 (2014), pages 5040–5053. https://doi.org/10.1109/TIT.2014.2323359.

[GM17] Danilo Garcia and Shane Macdonald. "Dark Personality Profiles: Estimating the Cluster Structure of the Dark Triad". In: *PsyCh Journal* (2017). https://doi.org/10.1002/pchj.175. https://www.researchgate.net/publication/316587259_Dark_Personality_Profiles_Estimating_the_Cluster_Structure_of_the_Dark_Triad.

[Gol] Andrew Goldstone. *tvtropes.org, the Future of the Humanities*. Blog post. https://arcade.stanford.edu/blogs/tvtropesorg-future-humanities.

[Gra73] Patrick Grant. "Tolkien: Archetype and Word". In: *CrossCurrents* 22 (1973). The JSTOR version did not seem to be free, so I read the maybe bootleg copy at the URL, pages 365–380. https://www.cgjungpage.org/learn/articles/book-reviews/35-tolkien-archetype-and-word.

[Hor+16] Fabian Horn, Cilliers Breytenbach, Camilla Di Biase-Dyson, Markus Egg, Therese Fuhrer, Verena Lobsien, Renate Schlesier, Jan Stenger, and Beatrice Trînca. "Spatial Metaphors of the Ancient World: Theory and Practice". In: *eTopoi Journal for Ancient Studies* Special Volume 6 (2016): Space and Knowledge (2016). ed. by Gerd Graßhoff and Michael Meyer, pages 453–480. ISSN: 2192-2608. http://www.topoi.org/wp-content/uploads/2016/12/eTopoi_Sp6_Horn-et-al.pdf.

[Kai13] David Kaiser. "Déjà vu all over again? A response to Philip Mirowski". In: *SERRC: Social Epistemology Review and Reply Collective* 2 (2013), pages 1–7. https://social-epistemology.com/2013/01/09/deja-vu-all-over-again-a-response-to-philip-mirowski-david-kaiser/.

[Kub+21] Emily Kubin, Curtis Puryear, Chelsea Schein, and Kurt Gray. "Personal experiences bridge moral and political divides better than facts". In: *Proceedings of the National Academy of Sciences* 118.6 (2021). ISSN: 0027-8424. https://doi.org/10.1073/pnas.2008389118. eprint: https://www.pnas.org/content/118/6/e2008389118.pdf. https://www.pnas.org/content/118/6/e2008389118.

[L2K21] O-Joun Lee, Eun-Soon You 2, and Jin-Taek Kim. "Plot Structure Decomposition in Narrative Multimedia by Analyzing Personalities of Fictional Characters". In: *Appl. Sci.* (2021). https://doi.org/10.3390/app11041645.

[LEn07] Madeleine L'Engle. *A Wind in the Door*. 2007. ISBN: 978-0-312-36854-8.

[LJ08] George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 2008. ISBN: 9780226470993.

[Mau13] Anna-Sofia Maurin. *Tropes*. 2013. https://plato.stanford.edu/entries/tropes/.

[McA93] Dan P. McAdams. *The Stories We Live By*. Guilford, 1993. ISBN: 978-1572301887.

[MR74] Albert Mehrabian and James. A Russell. "The Basic Emotional Impact of Environments". In: *Perceptual and Motor Skills* 38 (1974), pages 283–301.

[Oor21] Colin Van Oort. Personal correspondence. Apr. 28, 2021.

[PRL20] F Pianzola, S Rebora, and G Lauer. "Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins". In: *PLoS ONE* 15(1): e0226708 (2020). https://doi.org/10.1371/journal.pone.0226708.

[Rea+16] Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. "The emotional arcs of stories are dominated by six basic shapes". In: *EPJ Data Sci. 5, 31* (2016). https://doi.org/10.1140/epjds/s13688-016-0093-1. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0093-1#citeas.

[Rus80] James. A Russell. "A Circumplex Model of Affect". In: *Journal of Personality and Social Psychology* 39 (1980), pages 1161–1178.

[SCH17] Robyn Speer, Joshua Chin, and Catherine Havasi. *ConceptNet 5.5*. 2017. https://conceptnet.io.

[Wt97] Ludwig Wittgenstein and G.E.M. Anscombe (translator). *Philosophical investigations*. Originally published in German in 1953. Blackwell, 1997.

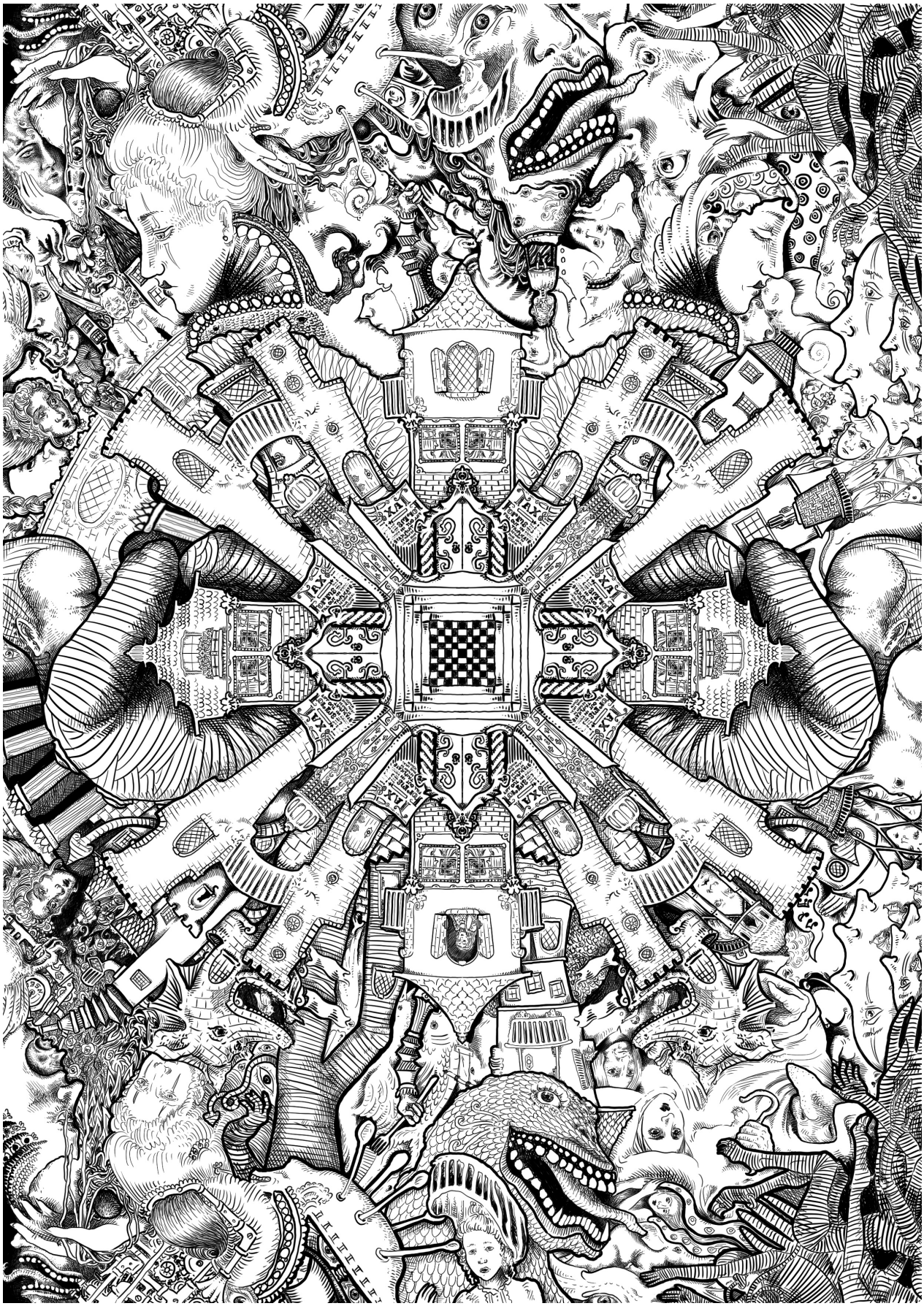

Figure 26: Visual interpretation of a lawful-chaotic manifold.