

Review



Cite this article: Perc M. 2020 Beauty in artistic expressions through the eyes of networks and physics. *J. R. Soc. Interface* **17**: 20190686.

<http://dx.doi.org/10.1098/rsif.2019.0686>

Received: 3 October 2019

Accepted: 17 February 2020

Subject Category:

Life Sciences—Physics interface

Subject Areas:

biocomplexity, evolution

Keywords:

complexity, entropy, network science, data science, self-organization

Author for correspondence:

Matjaž Perc

e-mail: matjaz.perc@gmail.com

Beauty in artistic expressions through the eyes of networks and physics

Matjaž Perc^{1,2,3}

¹Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia

²Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

³Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria

MP, 0000-0002-3087-541X

Beauty is subjective, and as such it, of course, cannot be defined in absolute terms. But we all know or feel when something is beautiful to us personally. And in such instances, methods of statistical physics and network science can be used to quantify and to better understand what it is that evokes that pleasant feeling, be it when reading a book or looking at a painting. Indeed, recent large-scale explorations of digital data have lifted the veil on many aspects of our artistic expressions that would remain forever hidden in smaller samples. From the determination of complexity and entropy of art paintings to the creation of the flavour network and the principles of food pairing, fascinating research at the interface of art, physics and network science abounds. We here review the existing literature, focusing in particular on culinary, visual, musical and literary arts. We also touch upon cultural history and culturomics, as well as on the connections between physics and the social sciences in general. The review shows that the synergies between these fields yield highly entertaining results that can often be enjoyed by layman and experts alike. In addition to its wider appeal, the reviewed research also has many applications, ranging from improved recommendation to the detection of plagiarism.

1. Introduction

The past decade has seen data science emerge as the new buzzword in the world of research. In 2012, Thomas H. Davenport and D. J. Patil at *Harvard Business Review* called being a data scientist ‘the sexiest job of the 21st century’. And although some argue that data science is not much more than classical statistics with a touch of modern flair, the richness of digital data today, in terms of volume, diversity and speed of change [1], requires synergies that transcend disciplinary boundaries. Data science uses methods and techniques from mathematics, physics, computer science, information science and of course statistics to make sense of the vast amounts of digital data that are daily added to computer servers worldwide. Visualization wizardry is likewise key for the data science to shine through, as opposed to rather one-dimensional and often dull representations one often finds in hardcore statistics books.

And it is this deluge of digital data that is often the bridge between the social and natural sciences, and, as we will review in what follows, also between art and natural sciences, and between art and physics and network science in particular. One may wonder whether classical physics has anything to do with societal challenges and with modern human societies. Yet research has shown that collectively we often behave no differently than particles in matter [2]. Not exactly, of course, but close enough for methods of statistical physics to be applied prolifically to subjects such as traffic [3], crime [4], epidemics [5], vaccination [6], cooperation [7], climate inaction [8], as well as antibiotic overuse [9] and moral behaviour [10]. The physics of social systems, or social physics [11] or sociophysics [12]—regardless of the name tag—has had

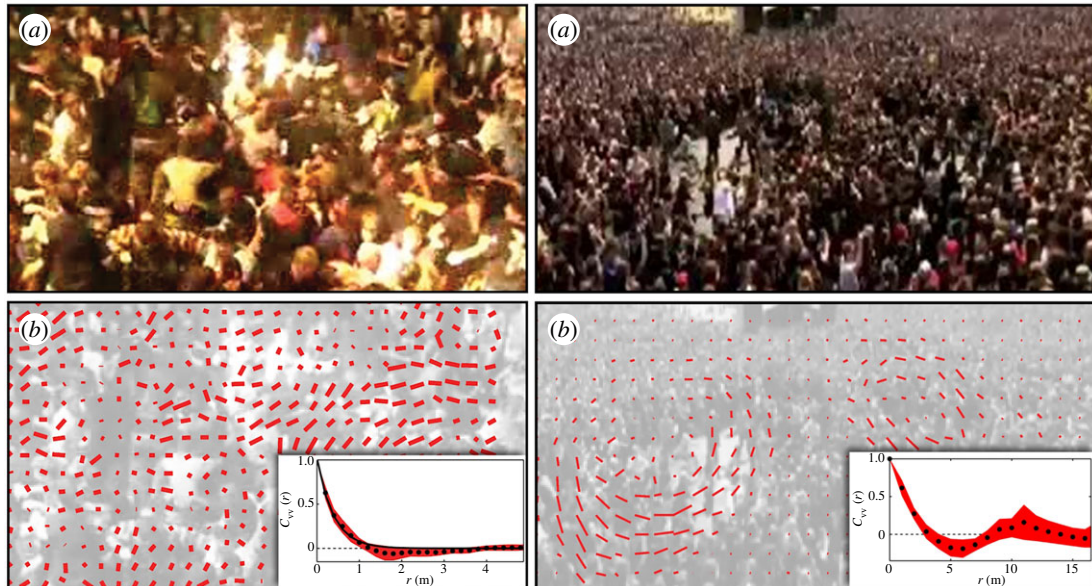


Figure 1. Collective human motion at heavy metal concerts. Left a mosh pit, akin to a disordered gas-like state. Right a circle pit, akin to an ordered vortex-like state. Both mosh and circle pits can be reproduced in flocking simulations (bottom row), demonstrating that human collective behaviour is consistent with the predictions of simplified models. For details concerning the mathematical model and the interpretation of insets in the bottom row with refer to the original work. Reproduced from [25].

an excellent growth pattern over the past couple of decades, spurred on by advances in statistical and computational physics, the availability of data, and the coming of age of related fields of research such as network science [13–16] and computational social science [17].

Looking at this development historically, it is, in fact, safe to claim that synergies between physical and social sciences have been around for centuries. Surely not as precise and useful as they are now, but nevertheless. Already in the seventeenth century, Thomas Hobbes based his theory of the state on the laws of motion, in particular, on the principle of inertia, which was then deduced by his contemporary Galileo Galilei. The ‘invisible hand’ proposed by Adam Smith in the second half of the eighteenth century is also eerily similar to the now famous notions of economic and social self-organization [18,19], and at the time was deemed to be as dependable in operation as the law of gravity [20]. And in the nineteenth century, the evolving physical theories of matter as a vast collection of atoms and molecules inspired a statistical view of societies and the predictable averages therein. French political and economic theorist Henri de Saint-Simon indeed proposed that society could be described by laws similar to those in physics. Just as the random movements of molecules in a gas yield the mathematically simple gas laws, it was fathomed that so might societies be predictable in the collective. Thus, as Philip Ball argued aptly [21], early sociology might have been constructed according to an unspoken faith that there was a kind of ‘physics of society’ that exists but is as yet somewhat in the stars.

But it is one thing to investigate and mathematically describe measurable phenomena in human societies, and a whole other to try and do the same for art. There is definitely something to art that will remain forever untouchable to the coldness of measure and science. But in spite of knowing full well we will likely never be able to understand it all, many have tried and succeeded in bridging the divide, with often deeply satisfying outcomes for both science and art. An example of this line of research is computational aesthetics

[22], having roots already in the first half of the twentieth century, when the American mathematician George D. Birkhoff proposed the ratio between order and complexity as an aesthetic measure [23]. In fact, the key mission of computational aesthetics is to develop scientific means to quantify beauty and model human aesthetic perception [24], although it influences also computer-generated art.

As something of a middle ground, one can come across research on human collective behaviour that is due to art, more specifically due to heavy metal music. Silverberg *et al.* [25] studied the collective motion of humans at heavy metal concerts, showing that such social gatherings generate extreme behaviours, ranging from a disordered gas-like state to an ordered vortex-like state, as shown in figure 1. The differences between such so-called mosh and circle pits can be reproduced in flocking simulations, thus confirming that even relatively very simple mathematical models can accurately describe the essence of human collective behaviour.

In what follows, we review research dedicated to culinary arts, visual arts and musical arts, and where appropriate we also touch upon cultural history [26,27] and culturomics [28], and we describe the connections between physics and the social sciences. We conclude with a discussion of the reviewed research and an outlook for future research.

2. Culinary arts

It is somewhat debatable whether cooking is art. Putting together a hamburger at a McDonald’s is more automation than anything, and it is certainly no art. But creating a delicious meal for your loved ones can be art, and culinary arts, in particular, have to do with the preparation and cooking of foods. But what does this have to do with physics and networks?

Perhaps not much on first glance, yet this is deceiving. About eight years ago Ahn *et al.* [29] published a paper where they introduced the flavour network to uncover fundamental principles of food pairing. The backbone of the flavour

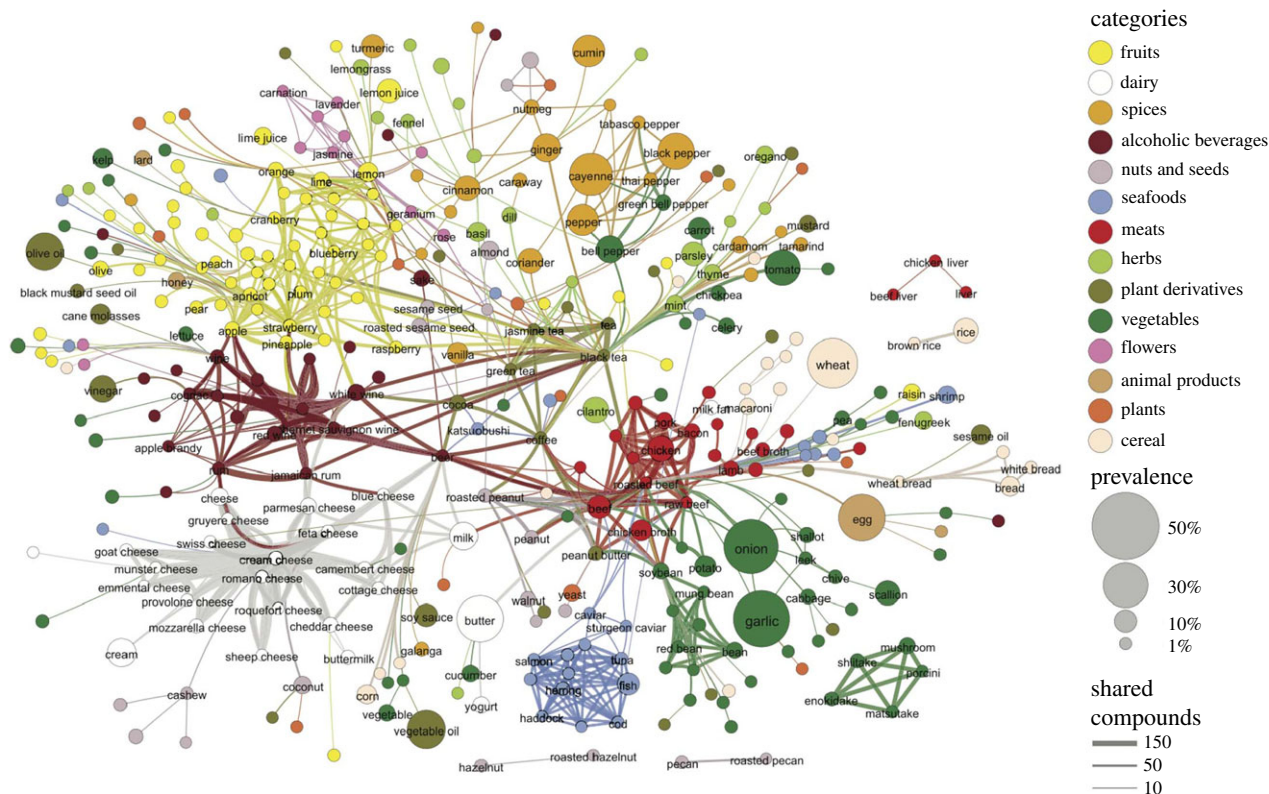


Figure 2. The flavour network. Each node represents an ingredient, the colour of nodes indicates food categories listed top right, and the node size, as show bottom right, reflects the prevalence of an ingredient in recipes. Two ingredients are connected if they share a significant number of flavour compounds, while the thickness of links represents the number of shared compounds between the two ingredients. Only statistically significant links at p -value cut-off 0.04 are show as otherwise the network would be too dense. For further details, we refer to the original work. Reproduced from [29].

network is shown in figure 2. They have argued that, given the increasing availability of information on food preparation, such as at cookpad.com and foodpairing.com, a data-driven approach can open up new avenues towards a systematic understanding of culinary art. Their research showed that North American and Western European cuisines have a statistically significant tendency towards sharing flavour compounds in recipe ingredients. By contrast, East Asian and Southern European recipes are much less likely to have ingredients that share flavour compounds. And vice versa, the more compounds are shared by two ingredients, the more likely they appear together in North American recipes, and less likely to appear together in East Asian cuisine. In exploring the mechanism responsible for these differences, Ahn *et al.* [29] found that the food-pairing effect is due to only a couple of ingredients that are frequently used in a particular cuisine. This would be milk, butter, cocoa, vanilla, cream and egg in North America, and beef, ginger, pork, cayenne, chicken and onion in East Asia, for example. These findings resonates with the well-known 'flavour principle' [30], according to which regional cuisines are due to just a few key ingredients, such as soy sauce for Asia or paprika and onion for Hungary.

However, already prior to the flavour network study [29], Kinouchi *et al.* [31] studied the statistics of ingredients and recipes taken from the Brazilian *Dona Benta*, the British *New Penguin Cookery Book*, the French *Larousse Gastronomique* and the medieval *Pleyn Delit*. They have observed universal distributions of ingredients with scale-invariant behaviour, which motivated a mathematical model akin to growth and preferential attachment in networks [32], or more generally to the Matthew effect [33]. The so-called copy-mutate model of culinary evolution was shown to fit the empirical data well.

The authors also argued that the model indicates an evolutionary dynamics governing the evolution of recipes over the centuries where idiosyncratic ingredients are preserved in a manner similar to the founder effect in biology [34].

In more recent years, several research works followed up to study culinary arts with methods of network science, physics and related approaches. For example, Teng *et al.* [35] showed how to improve recipe recommendations based on ingredient networks. Their research showed that recipe ratings can be predicted well with features from ingredient networks and nutritional information. The food-bridging hypothesis was also proposed, assuming that if two ingredients do not share a strong molecular or empirical affinity, they may become affine through a chain of pairwise affinities [36]. Together with the food-pairing hypothesis [29], four classes of cuisine can then be distinguished. Namely, East Asian cuisine that tends to avoid food-pairing and food-bridging, Latin American cuisine that tends to embrace food-pairing and food-bridging, South-eastern Asian cuisine that tends to avoid food-pairing but embraces food-bridging and Western cuisine that embraces food-pairing but avoids food-bridging [36].

Relatively, smaller-scale research efforts have also been made, for example, looking specifically at Arab cuisine and whether it abides to the food-pairing hypothesis [37]. Flavour pairing was also studied for the medieval European cuisine [38], where authors investigated the flavour pairing hypothesis historically, focusing in particular also on the role of data incompleteness and error. To that effect, two separate chemical compound datasets with different levels of cleanliness were used, showing that they give conflicting conclusions with regards to the flavour pairing hypothesis in medieval Europe. Possible inferences about the evolution of culinary

arts when many new ingredients are suddenly made available were also presented and discussed. The geography and similarity of regional cuisines in China has also been studied [39]. Research showed that geographical proximity rather than climate similarity is the crucial factor that determines the similarity of regional cuisines in China. Following the Kinouchi *et al.* [31] copy-mutate model of culinary evolution, similar models have also been proposed specifically for different Indian cuisines [40]. Authors argued that, in addition to the identified similarities and differences between the regions, a comparison of these models at the level of flavour compounds might open up a path towards molecular level studies that would associate specific ingredients with non-communicable diseases such as diabetes.

Apart from geographical interests and periodization, research at the interface of culinary arts and the more harder sciences also produced gems such as ‘Tell me what you eat, and I will tell you where you come from: A data science approach for global recipe data on the web’ [41] where the message is squarely in the title. It also gave rise to food computing [42], which acquires and analyses heterogeneous food data from disparate sources for perception, recognition, retrieval, recommendation and monitoring of food. Computational approaches can then be applied to address food-related issues in medicine, biology, gastronomy and agronomy. Online food preferences have also been explored based on server log data from a large recipe platform on the web [43]. Research revealed that recipe preferences are partly driven by ingredients, that recipe preference distributions exhibit more regional differences than ingredient preference distributions, and that weekday preferences are clearly distinct from weekend preferences. Along similar lines, food consumption and dietary patterns were also studied through twitter [44], aptly concluding ‘you tweet what you eat’, as well as through web usage logs [45].

We conclude this section with a reference to a contemporary book titled *Everyone Eats* [46], which explores why we eat what we eat, why some love spices, sweets and coffee, and why rice become such a staple food throughout so much of eastern Asia. With a focus on the social and cultural reasons for our food choices, the book may be a nice companion for further exploring this fascinating subject beyond physics and networks.

3. Visual arts

Patterns in nature are common and often beautiful and intriguing [47]. Pattern formation is thus, expectedly, also common in physics, chemistry and biology research, where emergent ordered and disordered structures often require quantification [48–54]. Indeed, the subject has been vibrant and popular ever since the seminal research by Alan Turing on the chemical basis of morphogenesis [55]. Perhaps thus not surprisingly, many methods that are often used in physics to quantify and study patterns can also be used to study art with hardly any modification.

The challenge up until recently was how to get visual art into quality digital form, and how to do so for a large number of artworks. With wikiart.org this challenge has been beautifully resolved, which opened up the path to large-scale analysis with methods of physics. Using these data, Sigaki *et al.* [56] studied the history of art paintings through the

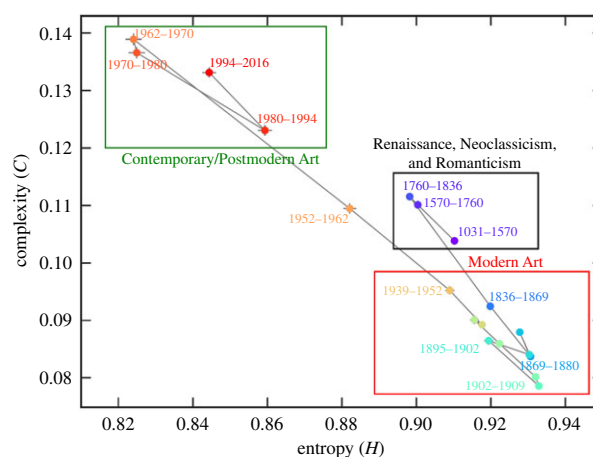


Figure 3. Quantifying the evolution of artworks through the history of art. Depicted is the temporal evolution of the average values of permutation entropy H and statistical complexity C . Each dot in the complexity–entropy plane corresponds to the average values of H and C for a given time interval, as indicated in the plot. Error bars represent the standard error of the mean. The highlighted regions show different art periods (black: Renaissance, Neoclassicism and Romanticism; red: Modern Art; green: Contemporary/Postmodern Art). We observe that the complexity–entropy plane correctly identifies different art periods and the transitions among them. Reproduced from [56].

lens of entropy and complexity. Almost 140 thousand paintings, spanning nearly a millennium of art history, have been included in the research. Interestingly, it was shown that the complexity–entropy plane reflects traditional concepts in art history, such as Wölfflin’s dual concepts of linear versus painterly and Riegl’s dichotomy of haptic versus optic [57,58]. Linear artworks are composed of clear and outlined shapes, while painterly artworks have contours that are subtle and smudged for merging image parts and passing the idea of fluidity. Similarly, haptic artworks depict objects as tangible discrete entities, isolated and circumscribed, whereas optic artworks represent objects as interrelated in deep space by exploiting light, colour and shadow effects to create the idea of an open spatial continuum. Relating this to complexity and entropy, it was shown that linear/haptic artworks are described by small values of entropy H and large values of complexity C , while painterly/optic artworks are expected to yield larger values of H and smaller values of C .

These facts then allowed Sigaki *et al.* [56] to quantify the evolution of artworks through the history of art. To do so, the average values of H and C after grouping the images by date were computed. Results shown in figure 3 reveal that the artworks produced between the ninth and the seventeenth century are on average more regular/ordered than those created between the nineteenth and the mid-twentieth century. Also, the artworks produced after 1950 are even more regular/ordered than those from the two earlier periods. It can be observed further that the pace of changes in the complexity–entropy plane intensifies after the nineteenth century. Notably, this period indeed coincides with the emergence of several artistic styles such as Neoclassicism and Impressionism. Moreover, the three outlined regions correspond well with the main divisions of art history, as indicated in figure 3.

Going a step further, complexity and entropy can also be used to distinguish different artistic styles in the H – C plane [56]. Since the values of H and C capture the degree of similarity among artistic styles regarding the local ordering of

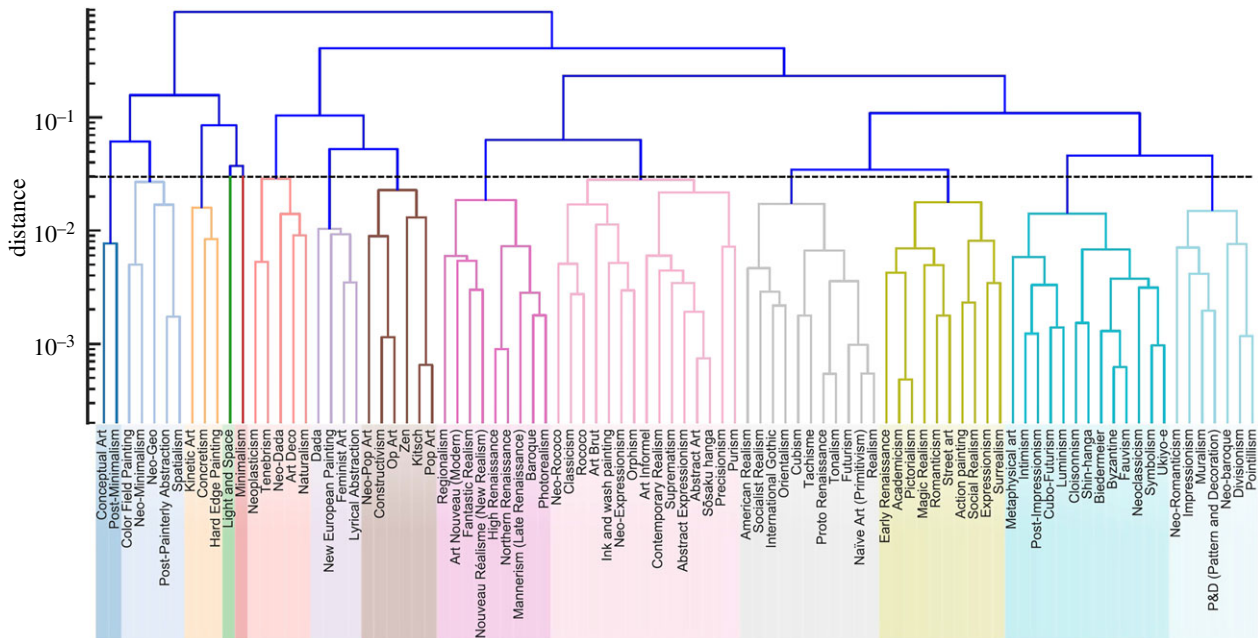


Figure 4. Hierarchical organization of artistic styles. Dendrogram representation of the distance matrix obtained by applying the minimum variance method. The 14 groups of styles indicated by the colourful branches are obtained by cutting the dendrogram at the threshold distance 0.03. This value maximizes the silhouette coefficient, and it is thus a natural number for defining the number of clusters in the dataset. For further details, we refer to the original work. Reproduced from [56].

image pixels, a test for a possible hierarchical organization of styles with respect to this local ordering is thus possible. To do so, the Euclidean distance between a pair of styles in the complexity–entropy plane can be considered as a dissimilarity measure between them. Thus, the closer the distance between two artistic styles, the more significant their similarity, and vice versa. Results obtained with this procedure are shown in figure 4, which can be interpreted as a tree of art, akin to the tree of life in biology, as per the metaphor made famous by Charles Darwin in his book *On the Origin of Species*.

Of course, the study by Sigaki *et al.* [56] is neither the first nor the last along this line of research. Already four years earlier Kim *et al.* [59] conducted a large-scale quantitative analysis of painting arts with the aim to bridge the gap between art and science, focusing on the usage of individual colours, the variety of colours, and the roughness of the brightness. Their research revealed a difference in colour usage between classical paintings and photographs, and a significantly lower colour variety during the medieval period. An increment in the roughness exponent in painting techniques such as chiaroscuro and sfumato has also been reported, consistent with the historical circumstances of the period. The same group later also studied the heterogeneity in chromatic distances in images of the modern era [60], where they noted rightfully that aggregate statistics are a poor measure for the individuality of painters since the differences can stem from different painters, or from the same painters simply employing different styles. An insightful analysis from [60] is shown in figure 5, where representative individual painters are characterized in detail. Two distinct yet complementary aspects of stylistic individuality are considered, namely the evolution over their careers, and uniqueness relative to the dominant style of the period. Ultimately, this research provides a treasure of insights for an extraordinary expansion in creative diversity and individuality that defines the modern era.

In addition to the above-reviewed large-scale research attempts at quantifying visual art, it is important to note

that the idea itself can be traced as far back as 1933, when the American mathematician George D. Birkhoff published his book *Aesthetic Measure* [23], which also gave rise to computational aesthetics [22]. Therein, he proposes to use the ratio between the number of regularities found in an image and the number of elements in an image as a quantitative aesthetic measure. But first applications happened only at the turn of the twenty-first century, where the work of Taylor *et al.* [61] showed that Pollock's paintings are characterized by an increasing fractal dimension over the course of his artistic career. This subsequently inspired the application of fractal analysis to the authenticity of paintings [62–66], to the evolution of artists [67,68], to the statistical properties of paintings [69] and artists [70–72], to art movements [73] and to other visual expressions [74–76]. And we would be remiss not to mention that fractal art is a fascinating subject on its own right [77]. It is common in Islamic culture, for example on the main dome of the Selimiye Mosque in Turkey, and in Hindu Temples around the world. There is also innovative recent research focusing on the analysis of large-scale datasets of paintings and other visual art expressions by means of the estimation of their average brightness and saturation [78,79].

To end the section on visual arts, we note that in addition to art paintings, sculpture, ceramics, photography, video, filmmaking, design, crafts and architecture of course also fall under this category. Notably, filmmaking has been studied in terms of actor networks—where two actors are connected if they appeared together in a movie. One of the more recent attempts in this direction is the social network analysis of character interactions in the *Stargate* and *Star Trek* television series [80]. Research revealed that the character networks of both series have small-world properties, and that the underlying network structure of an episode can tell us something about the complexity of the storyline in that particular episode. Episode networks were found to be either closed networks, networks with bottlenecks that connect otherwise disconnected clusters, or a mixture of the two, which could be

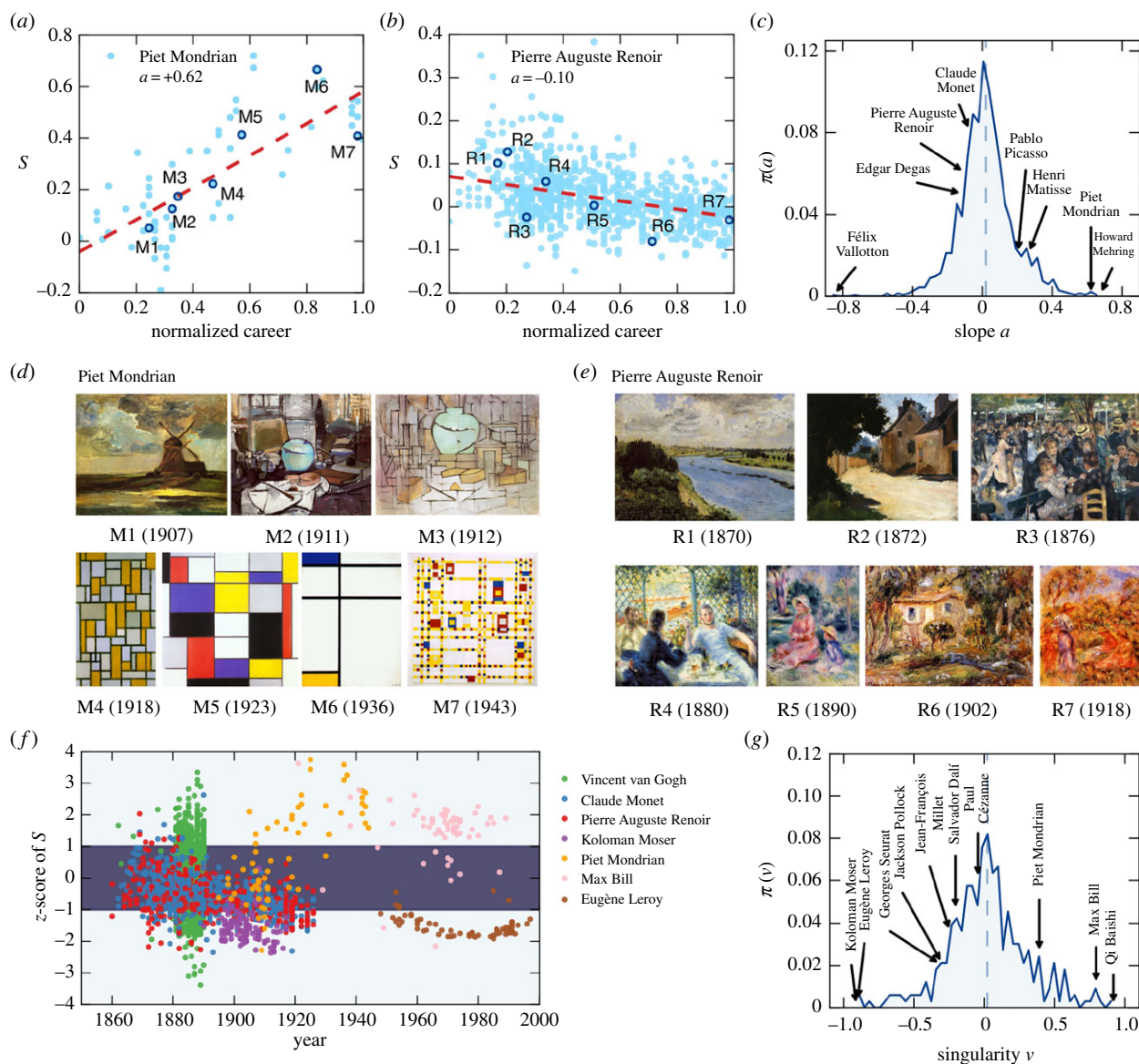


Figure 5. An in-depth analysis of individual painters. (a,b) The growth of the seamlessness in the paintings of Mondrian and Renoir over their career. Dashed red lines indicate best linear fits. In (c), a histogram of these linear fits is shown for 1326 artists that produced paintings in at least five distinct years. A few notable artists are indicated. (d,e) Samples of paintings by Mondrian and Renoir that highlight the stylistic changes of the two artists over their careers. (f) The singularity of paintings by seven select artists, while (g) shows the histogram of the singularity of 330 artists with more than 40 paintings. For further details, we refer to the original work. Reproduced from [60].

linked to the corresponding storylines upon a more detailed reading. But other forms of visual arts have yet to receive their link with physics and networks.

4. Musical arts

Research linking musical arts with science has the longest tradition amongst the arts considered in this review. Accordingly, we focus only on a selection of relatively recent research efforts that build predominantly on physics and network science as the bridges between the two disciplines.

We begin with Park *et al.* [81], who studied the topology and evolution of the network of western classical music composers. Based on the data from arkivmusic.com and allmusic.com, a bipartite network consisting of CDs and composers as the two node classes was created. Specifically, an edge between two nodes was drawn when a composer's pieces was recorded on the CD. A single-mode projection of just composers was also

made, where two composers were connected if they have been co-featured on a CD. Based on this, Park *et al.* [81] reported a wealth of results, including that the networks exhibit characteristics common to many real-world networks, such as the small-world property, the existence of a giant component, high clustering and heavy-tailed degree distributions. They have also explored the global association patterns of composers via centrality, assortative mixing and community structure, which suggest an intriguing interplay between the networks of musicians and our musicological understanding of the western musical tradition. Moreover, the study of the growth of the CD-composer bipartite network over time revealed superlinear preferential attachment as a strong candidate for explaining the increasing concentration of edges around top-degree nodes and the power-law degree distributions.

From this, we review results concerning the community structure, as shown in figure 6. Six sizable communities are depicted, which account for 99% of the 878 nodes with known periods. Fascinatingly, the communities roughly

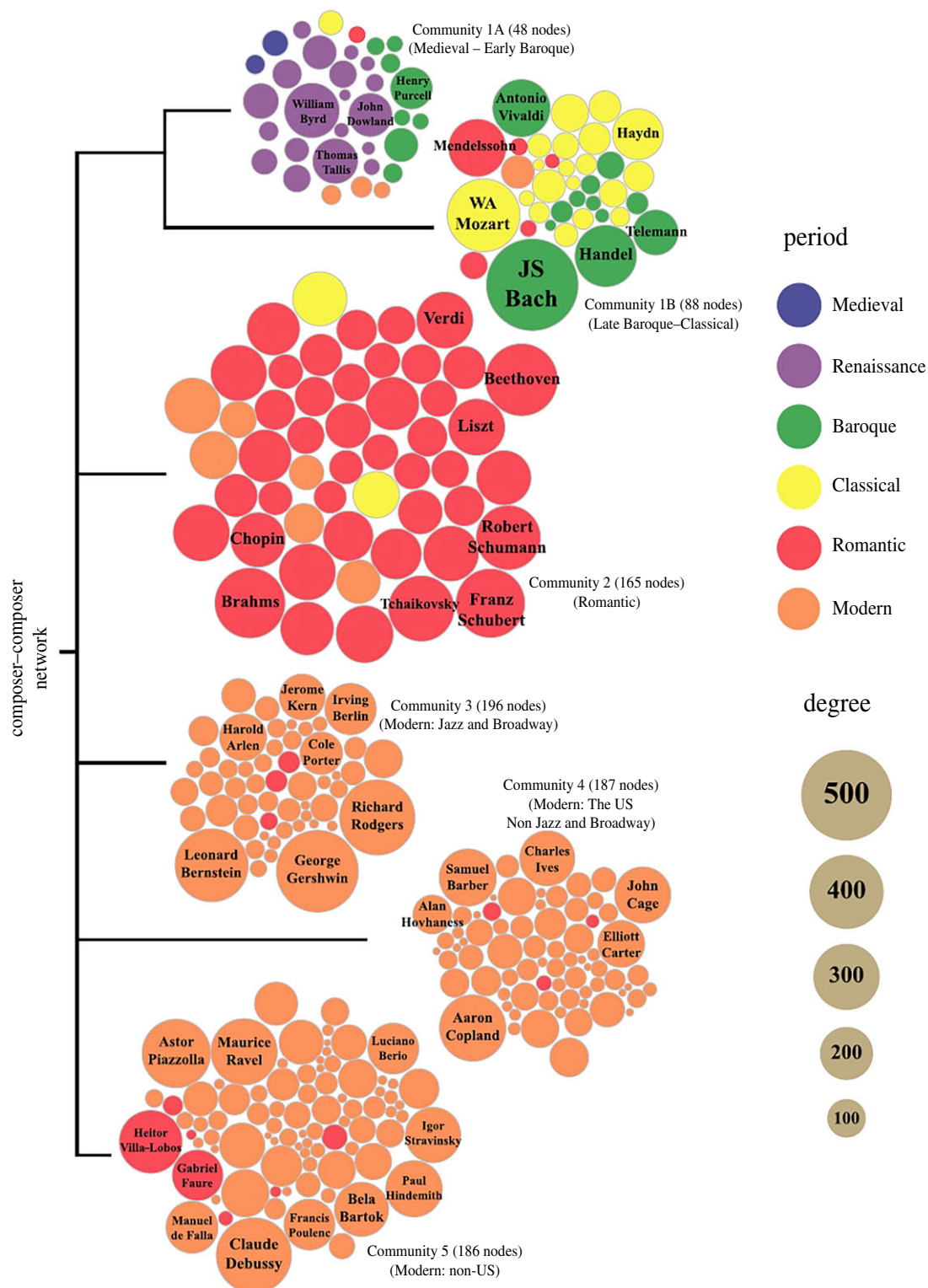


Figure 6. The community structure of the composer–composer network. Depicted are the five largest composer communities, which cover 6.2% of the composers who account for 60.1% of degrees. The communities correspond well to the established period definitions in classical music history. On the right is the colour code for the periods that are represented in each community, and the scale for their size. For details, we refer to the main text and the original work. Reproduced from [81].

correspond to the different musical periods, such as Renaissance and early Baroque (1A), late Baroque and Classical (1B), Romantic (2) and Modern (remaining three communities). The original work also lists notable composers in each of these communities, thus showcasing beautifully the power of network science to explore art in insightful ways.

A relatively recent classic in terms of a more physics-inspired approach, although still based on networks, is

the paper by Corrêa *et al.* [82], where authors explore the problem of automatic music genre classification by exploring rhythm-based features from a complex network representation. Specifically, a Markov model is proposed to analyse the temporal sequence of rhythmic notation events, and a feature analysis is performed by using the principal components analysis and the linear discriminant analysis. The former is an unsupervised multivariate statistical approach, while the

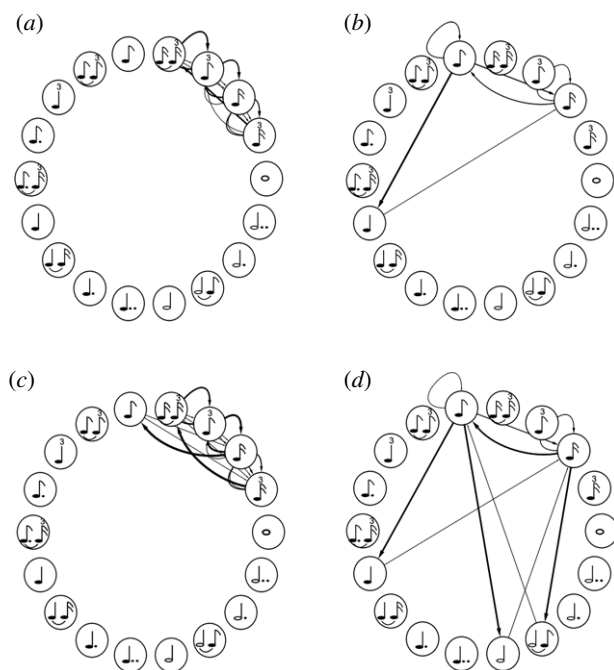


Figure 7. Digraph examples of four music samples as studied in [82]. Depicted are *How Blue Can You Get* by B. B. King (a), *Fotografia* by Tom Jobim (b), *Is This Love* by Bob Marley (c) and *From Me to You* by The Beatles (d). Reproduced from [82].

later is a supervised one. Two classifiers are also applied to identify the category of rhythms, namely the parametric Bayesian classifier under the Gaussian hypothesis and agglomerative hierarchical clustering. Figure 7 shows how the digraphs used for the analysis were created. It can be observed that only the durations of the notes, respecting the sequence in which they occur in the sample, were used to create the digraphs. Each vertex of the digraph represents one possible rhythm notation, such as a quarter note, a half note, an eighth note, and so on. The edges reflect the subsequent pairs of notes. For example, if there is an edge from vertex i , represented by a quarter note, to a vertex j , represented by an eighth note, this means that a quarter note was followed by an eighth note at least once. Moreover, the thicker the edges, the stronger the link between these two nodes.

Based on this approach, Corrêa *et al.* [82] have shown that the musical rhythms are surprisingly complex and contain many redundancies, and that many features are required to separate them, in fact too many to review here (please see original work). Research also revealed that by allowing multi-genre classification a generalization of the genre taxonomy can be achieved since new sub-genres emerge spontaneously from the original genres. Although the study focused only on the rhythmic analysis, authors noted that a deeper analysis of the rhythms can be performed to enhance the effectiveness of the methodology.

More recently, partly the same group of authors studied whether ten compositions by the Baroque composer Johann Sebastian Bach arose from a Markov chain [83], by using the recurrence quantification analysis. Perhaps not surprisingly, research found that the fairly implausible hypothesis that Bach's brain was a Markov chain can be consistently rejected with sufficiently elaborate surrogates.

Apart from the above-reviewed examples that showcase how different aspects of musical arts can be studied by means of networks and physics, research along similar lines

includes the application of complexity–entropy causality planes to distinguish songs [84], the identification of universal patterns in sound amplitudes of songs and music genres [85], the extraction of musical rhythmic patterns by means of community relevance in networks [86], and the quantification of soundscape dynamics of human agglomeration [87]. Paired with fundamental work on musical theory, for example concerning the geometry of musical chords [88], the hope is that such interdisciplinary research can contribute relevantly to the better understanding of music and its popularity across genres, geographical regions and time [89].

5. Literary arts

The digitalization of art had perhaps the biggest impact on literary arts. What previously existed only on paper became available as bits and bytes. Sites like the Google n-gram viewer at books.google.com/ngrams [28] and the Project Gutenberg at [gutenberg.org](https://www.gutenberg.org), as well as social networking sites such as Twitter and Facebook, strongly facilitated quantitative research inquiries into written text on massive scales [90–101]. However, the bulk of this research was concerned primarily with statistical properties, or with finding memes or with similarities or differences in the texts, rather than with the content in terms of art.

One fascinating research effort that did focus on content is due to Reagan *et al.* [102], who studied the emotional arcs for a filtered subset of 1327 stories from Project Gutenberg's fiction collection. As the main three independent methods of analysis, the authors used singular value decomposition—a standard linear algebra technique, Ward's method to generate a hierarchical clustering of stories, which proceeds by minimizing the variance between clusters of books [103], and the self-organizing map [104]—an unsupervised machine learning method to cluster emotional arcs. Emotional arcs were constructed by analysing the sentiment of sliding 10 000 word windows using [hedonometer.org](https://www.hedonometer.org) and the [labMT](https://www.labmt.com) dataset [105,106]. The outcome for the J. K. Rowling's *Harry Potter and the Deathly Hallows* is shown in figure 8, where the major highs and lows of the story are clearly inferable. As the authors note in their paper, the entire seven-book series can be classified as a 'Kill the monster' plot [107], while the many subplots and connections between them complicate the emotional arc of each individual book. The method, however, does not pick up emotional moments that are only discussed briefly, such as in a single paragraph or in a sentence. Whether the result shown in figure 8 corresponds with a reader's experience depends on many factors, but being proud owners of all the books and DVDs of the *Harry Potter* series in our family, it does seem to fit rather well. My teenage daughter Ela and her friends agree that the happiest should be the 'happy ever after' rather than the 'Harry at the Weasleys', but that is coming from a single group of people. Others may feel differently and agree fully with what is shown in figure 8.

In the grander picture, the main finding of the Reagan *et al.* [102] paper is that the emotional arcs of all the stories included in their research are dominated by no more than six basic shapes. These are 'Rags to riches' (rise), 'Tragedy' or 'Riches to rags' (fall), 'Man in a hole' (fall-rise), 'Icarus' (rise-fall), 'Cinderella' (rise-fall-rise) and 'Oedipus' (fall-rise-fall). Importantly, the same six emotional arcs are obtained from all possible arcs by observing them as the result of

Harry Potter and the Deathly Hallows

by J.K. Rowling

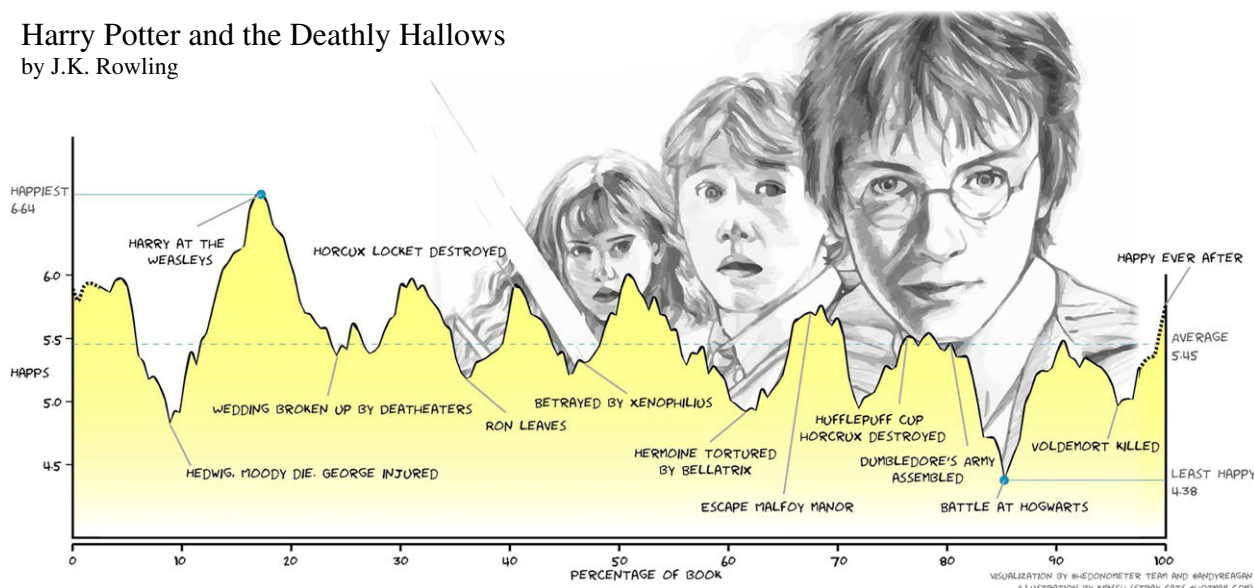


Figure 8. The emotional arc of *Harry Potter and the Deathly Hallows*, as determined by Reagan *et al.* [102], capturing the major highs and lows of the story. To generate such emotional arcs, authors analyse the sentiment of 10 000 word windows, which were slid through the text. The emotional content of each window was then evaluated using hedonometer.org with the labMT dataset [105,106]. The hedonometer.org webpage also provides interactive visualizations of emotional arc for many other books, stories, movie scripts, as well as for Twitter. Reproduced from [102].

three independent methods mentioned earlier, each with its own strength. Namely, the singular value decomposition finds the underlying basis of all of the emotional arcs, the clustering classifies the emotional arcs into distinct groups, and the self-organizing map generates arcs from noise which are similar to those in the corpus using a stochastic process. The result is thus truly robust and thoroughly supported by the presented evidence.

On a much smaller scale, Markovič *et al.* [108] recently studied the structure and complexity of texts in Slovene belles-lettres, with an emphasis on evaluating the differences in the texts for different age groups. Research revealed that the syntactic connectivity of words forms complex and heterogeneous networks that are characterized by an efficient transfer of information. It was also shown that with the increasing recommended age of readers, the length of texts, the average length of words, different combinations of phrases, and the complexity of social interactions between literary characters all increase. Conversely, the fraction of unique words was shown to decrease.

In terms of the language networks, as shown in figure 9, Markovič *et al.* [108] found that despite noticeable differences in network sizes, the degree distribution is a power law for all four age groups, and this with similar exponents. Apparently, some properties of syntactic patterns do not differ much for different age groups, although the relatively small sizes of the investigated networks did not allow a more precise comparison. Noteworthy, the observed scale-free property of language networks has been previously reported in several other works [110]. Further in terms of network properties, research revealed that the average degree and the average clustering coefficient increase with the recommended age of readers. This result goes in parallel with the decreasing tendency of unique word density. Namely, for higher age groups proportionally less unique words are used for longer texts and hence more word combinations are possible and indeed present, which in turn leads to more connections between individual words as well as to higher levels of

cliquishness. For the same reason, the networks also become less modular with the recommended age of readers. On account of the increasing number of connections, the average shortest path length progressively decreases with increasing recommended age, despite the huge increase in network size. This, in turn, gives rise to the small-world topological features of the extracted syntactic networks. Interestingly, despite changes in the average connectivity, the average shortest path length and network size, the diameter of the network remains largely unchanged across age groups. Taken together, network science enabled an in-depth theoretical exploration of Slovene belles-lettres, with clear distinctions in statistical properties between different age groups, thus bridging art and exact sciences in a mutually rewarding way.

In moving the subject further ahead, Ferraz de Arruda *et al.* [111] recently noted that, while well-established word-adjacency or co-occurrence networks successfully grasp syntactical features of written texts, they are unable to capture topical structure. To remedy this, they have proposed a network model wherein adjacent paragraphs act as nodes, which are then connected whenever they share a minimal semantic content. As an example, Lewis Carroll's *Alice's Adventures in Wonderland* has been studied, and research showed that such an approach can reveal many semantic traits of a text that would remain hidden otherwise. Moving away from 'just words' and phrases to different degrees of mesoscopic structure, such as sentences, paragraphs, or chapters, could be the next step in bringing literary art closer still to methods of network science and physics.

6. Discussion

We have reviewed recent research that aims to bridge the gap between different artistic expressions and network science and physics. Although most of the works that we have covered are not about beauty as such, in retrospect, they do allow us to

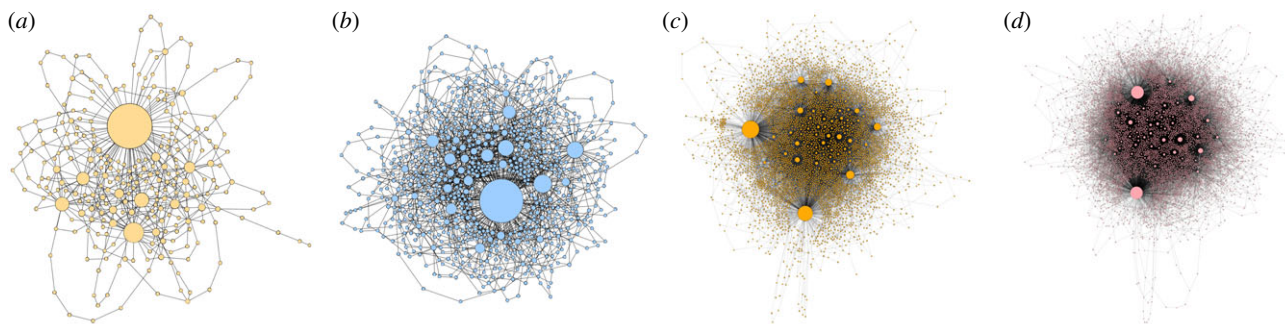


Figure 9. Language networks of selected Slovene belles-lettres for different age groups. Network A is for children age 1–5, network B is for children age 6–8, network C is for children age 9–11 and network D is for children age 12–14. Despite clearly visible differences in terms of the size and complexity of the networks, it is quite remarkable that the degree distribution is in all cases a power law. The average degree and the average clustering coefficient, however, increase with the recommended age of readers, while the modularity decreases. The average shortest path length decreases with increasing age, despite the increase in network size, which in turn gives rise to small-world properties [109]. Despite the changes in the average connectivity, the average shortest path length, and in the network size, the diameter of the networks is virtually identical in all age groups. For detailed data concerning these network properties, we refer to the original work. Reproduced from [108].

quantify and to understand what it is we find appealing or beautiful when we are subjected to a particular art form.

When it is food that we like, we can now appreciate which are the food pairings that make us take that extra bite, and which are the bridges between ingredients that make us revisit a restaurant. In broad terms, East Asian cuisine is our jam when neither food-pairing nor food-bridging should be on the menu, while Southeastern Asian cuisine appeals to those who do not like food-pairing but do like food-bridging. Western cuisine is all about food-pairing with almost no food-bridging, while Latin American cuisine should work well for those that like both food-pairing and food-bridging.

When we see a painting that we find beautiful, we can link this beauty to entropy and complexity. If we like ordered and repeating patterns, it is low entropy and high complexity, while everything ‘malerisch’ is high entropy and low complexity. The two physics quantities can be linked nicely to traditional concept in art history. Images formed by distinct and outlined parts yield many repetitions of a few ordinal patterns, and consequently, linear/haptic artworks are described by small values of entropy and large values of complexity. On the other hand, images composed of interrelated parts delimited by smudged edges produce more random patterns, and accordingly, painterly/optic artworks are expected to yield larger values of entropy and smaller values of complexity. It is possible to go even further and argue that Wölfflin’s dual concept of linear versus painterly and Riegl’s dichotomy of haptic versus optic are actually limiting forms of representation that demarcate the scale of all possibilities, and that in this regard the continuum of entropy and complexity values may help art historians to grade this scale more finely.

When we hear a song we then put on repeat, research at the interface of physics and network science can help us make out some of its properties that go beyond, or are outside the realm of, classic musical theory. Rhythm-based features from a network perspective, communities of different composers and genres, as well as complexity–entropy causality planes can all help us pinpoint what it is that we find beautiful in a particular song, improve personalized music recommendations, and improve also our understanding of what others around us find appealing in a particular piece of music. These approaches can also be used to try and forecast the future of several prominent artists and to decipher the growth dynamics of the music

network. Not to take anything away from classic musical theory and the geometry of musical chords and similar concepts, the point we are trying to make is simply that network science and physics can play a very useful role too.

Lastly, when we read a story that really moves us, we can look up whether its emotional arc is something like a ‘Cinderella’ rise-fall-rise, or perhaps just an ‘Icarus’ rise-fall. We can also appreciate whether we prefer many entangled characters that are almost hard to keep track of, or just two or three main characters that drive the story. A network science perspective will also give us insights about modularity of the story, that is whether some parts of it are particularly removed from other parts, and whether they come together at some point or not. Here, the concept of mesoscopic text structures, such as sentences, paragraphs or chapters, could be the new frontier in quantitative linguistics, not just in terms of statistics but also in terms of content and storytelling and emotional arcs.

The underlying movement of people, knowledge, craftsmanship and techniques that enabled the evolution of the arts reviewed above has, in a grander perspective, to do with our cultural history. In this regard, too methods of network science and physics have played an important role in recent years to advance the subject [26–28]. The culturomics paper by Michel *et al.* [28], the bravely titled Quantitative analysis of culture using millions of digitized books, was perhaps even too optimistic in what precisely we might be able to learn from such a ‘big data’ exploration of our culture. Later online comments and research were quite quick to point out the limits to inferences of socio-cultural and linguistic evolution based on such data [112], which consequently also meant that quite a few research efforts, such as [113,114] for example, might have been too quick to jump on the culturomics bandwagon [115]. But this is often the nature of a fast moving and growing field. The initial over-enthusiasm and positivity are dampened by more measured and careful approaches latter on, until eventually a maturity phase sets in, to be followed by the inevitable decline.

Nonetheless, the popularity of this line of research, together with connections to artificial intelligence via computational aesthetics [22], has been growing steadily in recent years, touching also upon science production [116–118]—see matjazperc.com/aps—and peer review [119] to complement research on many other aspects of our societies that we have mentioned

already in the introduction, including art as reviewed here. We hope this review will be informative and useful to researchers that are working at the fascinating interfaces of inherently different but also complementary research fields, and in particular to researchers that are seeking a way forward in mutually rewarding synergies between the arts and quantitative sciences.

References

- Leskovec J, Rajaraman A, Ullman JD. 2014 *Mining of massive datasets*. Cambridge, UK: Cambridge University Press.
- Castellano C, Fortunato S, Loreto V. 2009 Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646. (doi:10.1103/RevModPhys.81.591)
- Helbing D. 2001 Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067. (doi:10.1103/RevModPhys.73.1067)
- D'Orsogna MR, Perc M. 2015 Statistical physics of crime: a review. *Phys. Life Rev.* **12**, 1–21. (doi:10.1016/j.plev.2014.11.001)
- Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925. (doi:10.1103/RevModPhys.87.925)
- Wang Z, Bauch CT, Bhattacharyya S, Manfredi P, Perc M, Perra N, Salathé M, Zhao D. 2016 Statistical physics of vaccination. *Phys. Rep.* **664**, 1–113. (doi:10.1016/j.physrep.2016.10.006)
- Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. 2017 Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51. (doi:10.1016/j.physrep.2017.05.004)
- Pacheco JM, Vasconcelos VV, Santos FC. 2014 Climate change governance, cooperation and self-organization. *Phys. Life Rev.* **11**, 573–586. (doi:10.1016/j.plev.2014.02.003)
- Chen X, Fu F. 2018 Social learning of prescribing behavior can promote population optimum of antibiotic use. *Front. Phys.* **6**, 193. (doi:10.3389/fphy.2018.00139)
- Capraro V, Perc M. 2018 Grand challenges in social physics: in pursuit of moral behavior. *Front. Phys.* **6**, 107. (doi:10.3389/fphy.2018.00107)
- Perc M. 2019 The social physics collective. *Sci. Rep.* **9**, 16549. (doi:10.1038/s41598-019-53300-4)
- Galam S. 2004 Sociophysics: a personal testimony. *Physica A* **336**, 49–55. (doi:10.1016/j.physa.2004.01.009)
- Barrat A, Barthélemy M, Vespignani A. 2008 *Dynamical processes on complex networks*. Cambridge, UK: Cambridge University Press.
- Newman MEJ. 2010 *Networks: an introduction*. Oxford, UK: Oxford University Press.
- Estrada E. 2012 *The structure of complex networks: theory and applications*. Oxford, UK: Oxford University Press.
- Barabási A-L. 2015 *Network science*. Cambridge, UK: Cambridge University Press.
- Lazer D, Brewer D, Christakis N, Fowler J, King G. 2009 Life in the network: the coming age of computational social science. *Science* **323**, 721–723. (doi:10.1126/science.1167742)
- Mantegna RN, Stanley HE. 1999 *Introduction to econophysics: correlations and complexity in finance*. Cambridge, UK: Cambridge University Press.
- Helbing D. 2012 *Social self-organization*. Berlin, Germany: Springer.
- Smith A. 1759 *The theory of moral sentiments*. Strand & Edinburgh.
- Ball P. 2012 *Why society is a complex matter*. Berlin, Germany: Springer.
- Iqbal A. 2015 Computational aesthetics. Encyclopedia Britannica. (<https://www.britannica.com/topic/computational-aesthetics>)
- Birkhoff GD. 1933 *Aesthetic measure*. Cambridge, MA: Harvard University Press.
- Iqbal A, Van Der Heijden H, Guid M, Makhmalji A. 2012 Evaluating the aesthetics of endgame studies: a computational model of human aesthetic perception. *IEEE Trans. Comput. Intell. AI Games* **4**, 178–191. (doi:10.1109/TCIAIG.2012.2192933)
- Silverberg JL, Bierbaum M, Sethna JP, Cohen I. 2013 Collective motion of humans in Mosh and circle pits at heavy metal concerts. *Phys. Rev. Lett.* **110**, 228701. (doi:10.1103/PhysRevLett.110.228701)
- Schich M, Song C, Ahn YY, Mirsky A, Martino M, Barabási AL, Helbing D. 2014 A network framework of cultural history. *Science* **345**, 558–562. (doi:10.1126/science.1240064)
- Turchin P *et al.* 2018 Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc. Natl Acad. Sci. USA* **115**, E144–E151. (doi:10.1073/pnas.1708800115)
- Michel J-B *et al.* 2011 Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182. (doi:10.1126/science.1199644)
- Ahn Y-Y, Ahnert SE, Bagrow JP, Barabási A-L. 2011 Flavor network and the principles of food pairing. *Sci. Rep.* **1**, 196. (doi:10.1038/srep00196)
- Rozin E. 1971 *The flavor-principle cookbook*. Gloucestershire, UK: Hawthorn Books.
- Kinouchi O, Diez-García RW, Holanda AJ, Zambianchi P, Roque AC. 2008 The non-equilibrium nature of culinary evolution. *New J. Phys.* **10**, 073020. (doi:10.1088/1367-2630/10/7/073020)
- Barabási A-L, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
- Perc M. 2014 The Matthew effect in empirical data. *J. R. Soc. Interface* **11**, 20140378. (doi:10.1098/rsif.2014.0378)
- Mayr E. 1963 *Animal species and evolution*. Cambridge, MA: Harvard University Press.
- Teng C-Y, Lin Y-R, Adamic L. 2012 Recipe recommendation using ingredient networks. In *Proc. 4th Annual ACM Web Science Conf.*, pp. 298–307 ACM.
- Simas T, Ficek M, Diaz-Guilera A, Obrador P, Rodriguez PR. 2017 Food-bridging: a new network construction to unveil the principles of cooking. *Front. ICT* **4**, 14. (doi:10.3389/fict.2017.00014)
- Tallab ST, Alrazgan MS. 2016 Exploring the food pairing hypothesis in Arab cuisine: a study in computational gastronomy. *Procedia Comput. Sci.* **82**, 135–137. (doi:10.1016/j.procs.2016.04.020)
- Varshney KR, Varshney LR, Wang J, Myers D. 2013 Flavor pairing in medieval european cuisine: a study in cooking with dirty data. (<http://arxiv.org/abs/1307.7982>).
- Zhu YX, Huang J, Zhang ZK, Zhang QM, Zhou T, Ahn YY. 2013 Geography and similarity of regional cuisines in china. *PLoS ONE* **8**, e79161. (doi:10.1371/journal.pone.0079161)
- Jain A, Bagler G. 2018 Culinary evolution models for indian cuisines. *Physica A* **503**, 170–176. (doi:10.1016/j.physa.2018.02.176)
- Kim K-J, Chung C-H. 2016 Tell me what you eat, and I will tell you where you come from: a data science approach for global recipe data on the web. *IEEE Access* **4**, 8199–8211. (doi:10.1109/ACCESS.2016.2600699)
- Min W, Jiang S, Liu L, Rui Y, Jain R. 2019 A survey on food computing. *ACM Comput. Surv.* **52**, 92.
- Wagner C, Singer P, Strohmaier M. 2014 The nature and evolution of online food preferences. *Eur. Phys. J. Data Sci.* **3**, 38. (doi:10.1140/epjds/s13688-014-0036-7)
- Abbar S, Mejova Y, Weber I. 2015 You tweet what you eat: Studying food consumption through twitter. In *Proc. 33rd Annual ACM Conf. on Human Factors in Computing Systems*. pp. 3197–3206, ACM.
- West R, White RW, Horvitz E. 2013 From cookies to cooks: insights on dietary patterns via analysis of web usage logs. In *Proc. 22nd Int. Conf. on World Wide Web*. pp. 1399–1410, ACM.
- Anderson EN. 2014 *Everyone eats: understanding food and culture*. New York, NY: NYU Press.
- Ball P. 2016 *Patterns in nature*. Chicago, IL: The University of Chicago Press.

48. Harrison LG. 2005 *Kinetic theory of living pattern*. Cambridge, UK: Cambridge University Press.
49. Perc M. 2005 Spatial decoherence induced by small-world connectivity in excitable media. *New J. Phys.* **7**, 252. (doi:10.1088/1367-2630/7/1/252)
50. Sagués F, Sancho JM, García-Ojalvo J. 2007 Spatiotemporal order out of noise. *Rev. Mod. Phys.* **79**, 829. (doi:10.1103/RevModPhys.79.829)
51. Nakamasu A, Takahashi G, Kanbe A, Kondo S. 2009 Interactions between zebrafish pigment cells responsible for the generation of turing patterns. *Proc. Natl Acad. Sci. USA* **106**, 8429–8434. (doi:10.1073/pnas.0808622106)
52. Kondo S, Miura T. 2010 Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* **329**, 1616–1620. (doi:10.1126/science.1179047)
53. Szolnoki A, Perc M, Szabó G. 2012 Defense mechanisms of empathetic players in the spatial ultimatum game. *Phys. Rev. Lett.* **109**, 078701. (doi:10.1103/PhysRevLett.109.078701)
54. Andrade-Silva I, Bortolozzo U, Clerc MG, González-Cortés G, Residori S, Wilson M. 2018 Spontaneous light-induced turing patterns in a dye-doped twisted nematic layer. *Sci. Rep.* **8**, 12867. (doi:10.1038/s41598-018-31206-x)
55. Turing AM. 1952 The chemical basis of morphogenesis. *Phil. Trans. R. Soc. B* **237**, 37–72. (doi:10.1098/rstb.1952.0012)
56. Sigaki HYD, Perc M, Ribeiro HV. 2018 History of art paintings through the lens of entropy and complexity. *Proc. Natl Acad. Sci. USA* **115**, E8585–E8594. (doi:10.1073/pnas.1800083115)
57. Wölfflin H. 1950 *Principles of art history: the problem of the development of style in later art*. Mineola, NY: Dover.
58. Riegl A. 2004 *Historical grammar of the visual arts*. New York, NY: Zone Book.
59. Kim D, Son S-W, Jeong H. 2014 Large-scale quantitative analysis of painting arts. *Sci. Rep.* **4**, 7370. (doi:10.1038/srep07370)
60. Lee B, Kim D, Sun S, Jeong H, Park J. 2018 Heterogeneity in chromatic distance in images and characterization of massive painting data set. *PLoS ONE* **13**, e0204430. (doi:10.1371/journal.pone.0204430)
61. Taylor RP, Micolich AP, Jonas D. 1999 Fractal analysis of Pollock's drip paintings. *Nature* **399**, 422. (doi:10.1038/20833)
62. Jones-Smith K, Mathur H. 2006 Fractal analysis: revisiting Pollock's drip paintings. *Nature* **444**, E9–E10. (doi:10.1038/nature05398)
63. Taylor RP, Micolich AP, Jonas D. 2006 Fractal analysis: revisiting Pollock's drip paintings (reply). *Nature* **444**, E10–E11. (doi:10.1038/nature05399)
64. Taylor RP, Guzman R, Martin TP, Hall GDR, Micolich AP, Jonas D, Scannell BC, Fairbanks MS, Marlow CA. 2007 Authenticating Pollock paintings using fractal geometry. *Pattern Recognit. Lett.* **28**, 695–702. (doi:10.1016/j.patrec.2006.08.012)
65. Jones-Smith K, Mathur H, Krauss LM. 2009 Drip paintings and fractal analysis. *Phys. Rev. E* **79**, 046111. (doi:10.1103/PhysRevE.79.046111)
66. De la Calleja EM, Cervantes F, De la Calleja J. 2016 Order-fractal transitions in abstract paintings. *Annal. Phys.* **371**, 313–322. (doi:10.1016/j.aop.2016.04.007)
67. Boon JP, Casti J, Taylor RP. 2011 Artistic forms and complexity. *Nonlinear Dyn. Psychol Life Sci.* **15**, 265.
68. Alvarez-Ramirez J, Ibarra-Valdez C, Rodriguez E. 2016 Fractal analysis of Jackson Pollock's painting evolution. *Chaos Solitons Fractals* **83**, 97–104. (doi:10.1016/j.chaos.2015.11.034)
69. Pedram P, Jafari GR, Lisa M. 2008 The stochastic view and fractality in color space. *Int. J. Mod. Phys. C* **19**, 855–866. (doi:10.1142/S0129183108012558)
70. Taylor R. 2004 Pollock, Mondrian and the nature: recent scientific investigations. *Chaos Complex. Lett.* **1**, 29.
71. Hughes JM, Graham DJ, Rockmore DN. 2010 Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proc. Natl Acad. Sci. USA* **107**, 1279–1283. (doi:10.1073/pnas.0910530107)
72. Shamir L. 2012 Computer analysis reveals similarities between the artistic styles of Van Gogh and Pollock. *Leonardo* **45**, 149–154. (doi:10.1162/LEON_a_00281)
73. Elsa M, Zenit R. 2017 Topological invariants can be used to quantify complexity in abstract paintings. *Knowledge-Based Syst.* **126**, 48–55. (doi:10.1016/j.knsys.2017.03.030)
74. Castrejon-Pita JR, Castrejón-Pita AA, Sarmiento-Galán A, Castrejón-García R. 2003 Nasca lines: a mystery wrapped in an enigma. *Chaos* **13**, 836–838. (doi:10.1063/1.1587031)
75. Koch M, Denzler J, Redies C. 2010 $1/f^2$ Characteristics and isotropy in the Fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. *PLoS ONE* **5**, e12268. (doi:10.1371/journal.pone.0012268)
76. Montagner C, Linhares JMM, Vilarigues M, Nascimento SMC. 2016 Statistics of colors in paintings and natural scenes. *J. Opt. Soc. Am. A* **33**, A170–A177. (doi:10.1364/JOSAA.33.00A170)
77. Bovill C. 1996 *Fractal geometry in architecture and design*. Basel, Switzerland: Birkhauser.
78. Manovich L. 2015 Data science and digital art history. *Intl. J. Digital Art Hist.* **1**, 11–35. (doi:10.5007/1807-9288.2015v11n2p35)
79. Yazdani M, Chow J, Manovich L. 2017 Quantifying the development of user-generated art during 2001–2010. *PLoS ONE* **12**, e0175350. (doi:10.1371/journal.pone.0175350)
80. Tan MSA, Ujum EA, Ratnavelu K. 2017 Social network analysis of character interaction in the Stargate and Star Trek television series. *Int. J. Mod. Phys. C* **28**, 1750017. (doi:10.1142/S0129183117500176)
81. Park D, Bae A, Schich M, Park J. 2015 Topology and evolution of the network of western classical music composers. *EPJ Data Sci.* **4**, 2. (doi:10.1140/epjds/s13688-015-0039-z)
82. Corrêa DC, Saito JH, da F Costa L. 2010 Musical genres: beating to the rhythms of different drums. *New J. Phys.* **12**, 053030. (doi:10.1088/1367-2630/12/5/053030)
83. Moore JM, Corrêa DC, Small M. 2018 Is Bach's brain a Markov chain? Recurrence quantification to assess Markov order for short, symbolic, musical compositions. *Chaos* **28**, 085715. (doi:10.1063/1.5024814)
84. Ribeiro HV, Zunino L, Mendes RS, Lenzi EK. 2012 Complexity-entropy causality plane: a useful approach for distinguishing songs. *Physica A* **391**, 2421–2428. (doi:10.1016/j.physa.2011.12.009)
85. Mendes RS, Ribeiro HV, Freire F, Tateishi A, Lenzi EK. 2011 Universal patterns in sound amplitudes of songs and music genres. *Phys. Rev. E* **83**, 017101. (doi:10.1103/PhysRevE.83.017101)
86. Coca AE, Zhao L. 2016 Musical rhythmic pattern extraction using relevance of communities in networks. *Inf. Sci.* **329**, 819–848. (doi:10.1016/j.ins.2015.09.030)
87. Ribeiro HV, De Souza RT, Lenzi EK, Mendes RS, Evangelista LR. 2011 The soundscape dynamics of human agglomeration. *New J. Phys.* **13**, 023028. (doi:10.1088/1367-2630/13/2/023028)
88. Tymoczko D. 2006 The geometry of musical chords. *Science* **313**, 72–74. (doi:10.1126/science.1126287)
89. Miles SA, Rosen DS, Grzywacz NM. 2017 A statistical analysis of the relationship between harmonic surprise and preference in popular music. *Front. Hum. Neurosci.* **11**, 263. (doi:10.3389/fnhum.2017.00263)
90. Serrano MA, Flammini Á, Menczer F. 2009 Modeling statistical properties of written text. *PLoS ONE* **4**, e5372. (doi:10.1371/journal.pone.0005372)
91. Bernhardsson S, Correa da Rocha LE, Minnhagen P. 2009 The meta book and size-dependent properties of written language. *New J. Phys.* **11**, 123015. (doi:10.1088/1367-2630/11/12/123015)
92. Leskovec J, Backstrom L, Kleinberg J. 2009 Meme-tracking and the dynamics of the news cycle. In *Proc. ACM SIGKDD*, pp. 497–506.
93. Dodds PS, Danforth CM. 2010 Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *J. Happiness Stud.* **11**, 441–456. (doi:10.1007/s10902-009-9150-9)
94. Simmons MP, Adamic LA, Adar E. 2011 Memes online: extracted, subtracted, injected, and recollected. In *Proc. ICWSM*, 353–360.
95. Perc M. 2012 Evolution of the most common English words and phrases over the centuries. *J. R. Soc. Interface* **9**, 3323–3328. (doi:10.1098/rsif.2012.0491)
96. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. 2013 The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE* **8**, e64417. (doi:10.1371/journal.pone.0064417)
97. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. 2013 The Twitter of Babel: mapping world languages through microblogging platforms. *PLoS ONE* **8**, e61981. (doi:10.1371/journal.pone.0061981)
98. Gerlach M, Altmann EG. 2013 Stochastic model for the vocabulary growth in natural languages. *Phys. Rev. X* **3**, 021006. (doi:10.1103/PhysRevX.3.021006)
99. Dodds PS *et al.* 2015 Human language reveals a universal positivity bias. *Proc. Natl Acad. Sci. USA* **112**, 2389–2394. (doi:10.1073/pnas.1411678112)

100. Corral A, Boleda G, Ferrer-i Cancho R. 2015 Zipf's law for word frequencies: word forms versus lemmas in long texts. *PLoS ONE* **10**, e0129031. (doi:10.1371/journal.pone.0129031)
101. Bentz C, Alikaniotis D, Cysouw M, Ferrer-i Cancho R. 2017 The entropy of words-learnability and expressivity across more than 1000 languages. *Entropy* **19**, 275. (doi:10.3390/e19060275)
102. Reagan AJ, Mitchell L, Kiley D, Danforth CM, Dodds PS. 2016 The emotional arcs of stories are dominated by six basic shapes. *Eur. Phys. J. Data Sci.* **5**, 31. (doi:10.1140/epjds/s13688-016-0093-1)
103. Ward Jr JH. 1963 Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244. (doi:10.1080/01621459.1963.10500845)
104. Kohonen T. 1990 The self-organizing map. *Proc. IEEE* **78**, 1464–1480. (doi:10.1109/5.58325)
105. Reagan AJ, Danforth CM, Tivnan B, Williams JR, Dodds PS. 2017 Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *Eur. Phys. J. Data Sci.* **6**, 28. (doi:10.1140/epjds/s13688-017-0121-9)
106. Ribeiro FN, Araújo M, Gonçalves P, Gonçalves MA, Benevenuto F. 2016 SentiBench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *Eur. Phys. J. Data Sci.* **5**, 23. (doi:10.1140/epjds/s13688-016-0085-1)
107. Booker C. 2006 *The seven basic plots: why we tell stories*. New York, NY: Bloomsbury Academic.
108. Markovič R, Gosak M, Perc M, Marhl M, Grubelnik V. 2019 Applying network theory to fables: complexity in Slovene belles-lettres for different age groups. *J. Complex Netw.* **7**, 114–127. (doi:10.1093/comnet/cny018)
109. Watts DJ, Strogatz SH. 1998 Collective dynamics of 'small world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
110. Solé RV, Corominas-Murtra B, Valverde S, Steels L. 2010 Language networks: their structure, function, and evolution. *Complexity* **15**, 20–26. (doi:10.1002/cplx.20305)
111. Ferraz de Arruda H, Nascimento Silva F, Queiroz Marinho V, Raphael Amancio D, da Fontoura Costa L. 2017 Representation of texts as complex networks: a mesoscopic approach. *J. Complex Netw.* **6**, 125–144. (doi:10.1093/comnet/cnx023)
112. Pechenick EA, Danforth CM, Dodds PS. 2015 Characterizing the Google Books corpus: strong limits to inferences of socio-cultural and linguistic evolution. *PLoS ONE* **10**, e0137041. (doi:10.1371/journal.pone.0137041)
113. Gao J, Hu J, Mao X, Perc M. 2012 Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. *J. R. Soc. Interface* **9**, 1956–1964. (doi:10.1098/rsif.2011.0846)
114. Petersen AM, Tenenbaum JN, Havlin S, Stanley HE, Perc M. 2012 Languages cool as they expand: allometric scaling and the decreasing need for new words. *Sci. Rep.* **2**, 943. (doi:10.1038/srep00943)
115. Shannon CE. 1956 The bandwagon. *IRE Trans. Inf. Theory* **2**, 3. (doi:10.1109/TIT.1956.1056774)
116. Perc M. 2013 Self-organization of progress across the century of physics. *Sci. Rep.* **3**, 1720. (doi:10.1038/srep01720)
117. Kuhn T, Perc M, Helbing D. 2014 Inheritance patterns in citation networks reveal scientific memes. *Phys. Rev. X* **4**, 041036. (doi:10.1103/PhysRevX.4.041036)
118. Sinatra R, Wang D, Deville P, Song C, Barabási A-L. 2016 Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239. (doi:10.1126/science.aaf5239)
119. Ballestri S, Goldstone RL, Helbing D. 2016 Peer review and competition in the art exhibition game. *Proc. Natl Acad. Sci. USA* **113**, 8414–8419. (doi:10.1073/pnas.1603723113)