



OPEN

Entropy and complexity unveil the landscape of memes evolution

Carlo M. Valensise¹✉, Alessandra Serra², Alessandro Galeazzi⁶, Gabriele Etta⁵, Matteo Cinelli^{3,4} & Walter Quattrociocchi⁵

On the Internet, information circulates fast and widely, and the form of content adapts to comply with users' cognitive abilities. Memes are an emerging aspect of the internet system of signification, and their visual schemes evolve by adapting to a heterogeneous context. A fundamental question is whether they present culturally and temporally transcendent characteristics in their organizing principles. In this work, we study the evolution of 2 million visual memes published on Reddit over ten years, from 2011 to 2020, in terms of their statistical complexity and entropy. A combination of a deep neural network and a clustering algorithm is used to group memes according to the underlying templates. The grouping of memes is the cornerstone to trace the growth curve of these objects. We observe an exponential growth of the number of new created templates with a doubling time of approximately 6 months, and find that long-lasting templates are associated with strong early adoption. Notably, the creation of new memes is accompanied with an increased visual complexity of memes content, in a continuous effort to represent social trends and attitudes, that parallels a trend observed also in painting art.

Social media radically changed the way we consume information and interact online¹⁻³. Online interactions, indeed, influence social dynamics by favoring the formation of homophilic groups around shared narratives and attitudes and thus bursting group polarization⁴⁻⁶. In this scenario, multimedia content such as videos, photos, and pictures represents an essential portion of online communication, especially within social media platforms. Online communication can be read through the lenses of Dawkins' *cultural memes*⁷, whose definition applies to almost all online information vehicles. Cultural memes represent a unit of cultural information transmitted and replicated; writing posts, sharing personal videos, expressing "likes" are examples of this concept. While Dawkins' model of cultural evolution is nowadays considered insufficient to comprehend the complex cultural phenomena of information transmission⁸⁻¹¹, its evolutionary pattern still represents a valid basis for describing fundamental features of memes diffusion. In this work, we investigate the role and evolution of a particular kind of cultural meme, namely template images that undergo modifications or get some text overlapped, conventionally referred to just as *memes*. In the following, we adopt this convention. According to Dawkins' hypothesis, cultural memes¹² are characterized by the three essential elements of evolutionary theory: replication, variation, and selection. In the case of visual memes, the replication mechanism is self-evident. It consists of modifying an image, e.g., with some text, to represent a given situation. Moreover, replication of memes is facilitated by their consistency with other cultural memes present in the online environment, such as short videos, pictures, or short texts. Variation is an intrinsic feature of visual memes. Indeed, new memes are continuously created to target funny situations or jokes about political or societal events and compete for users' attention flowing across online communities. Finally, selection occurs when a meme cannot attract human attention nor adapt to transmit new contents and disappears adapting to the fast online environment. Among the online cultural memes that underwent relatively strong selection, we find, for example, blogs and discussion forums that have been replaced mainly by online social media; similarly, the emoji's introduction strongly reduced the use of *ascii art* symbols.

So far, a large body of research quantitatively investigated the features of different online cultural memes, not limited to images. Textual memes were analyzed by Leskovec et al.¹³ as a proxy for the cycle of online news consumption. Ienco et al.¹⁴ studied the problem of ranking memes, i.e., selecting those memes to be displayed to users to maximize the network activity on the platform. Romero et al.¹⁵ studied online memes propagation in the form of Twitter hashtags. Bauckhage¹⁶ investigated the epidemic dynamics of 150 famous memes, applying

¹Enrico Fermi Research Center, Piazza del Viminale, 1, 00184 Roma, Italy. ²Tuscia University - DISTU Department of Modern Languages and Literatures, History, Philosophy and Law Studies, Via S. Carlo, 32, 01100 Viterbo, Italy. ³Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari, University of Venice, Via Torino 155, 30172 Mestre, Italy. ⁴Institute for Complex Systems, Italian National Research Council, Via dei Taurini 19, 00185 Roma, Italy. ⁵Department of Computer Science, Sapienza University of Rome, Viale Regina Elena, 295, 00161 Roma, Italy. ⁶University of Brescia, Via Branze, 59, 25123 Brescia, Italy. ✉email: carlo.valensise@cref.it

models from mathematical epidemiology to account for the growth and decline of visual memes. Ratkiewicz and coworkers¹⁷ developed a framework for analyzing the diffusion of politics-related tweets. Weng et al.¹⁸ studied meme virality through an agent-based approach, accounting for the limited attention each user can spend in online environments. The mechanisms of competition and collaboration among memes, that determine the successful passage of a meme from a generation to another one has been investigated by Coscia¹⁹, together with the rules underlying memes popularity^{20,21}. These studies show that the dynamics behind memes (both visual and textual) virality are not linearly dependent upon the adherence to the template of new meme instances. The popularity of memes is investigated also with the underlying network community structure²². Ferrara et al.²³ employed clustering techniques to identify text-based memes, leveraging the content, metadata, and network structure of social data. Dang et al.²⁴ used tri-grams to cluster posts from Reddit and track the spreading of memes. Linguistic features²⁵ have been also investigated as predictors for textual memes popularity. Adamic et al.²⁶ explored a large corpus of textual data from Facebook modeling the propagation of information as a Yule process. Scientific memes²⁷ in the form of words were investigated exploring inheritance patterns in the scientific literature. Dubey et al.²⁸ employed a Deep Learning architecture to process memes, extract the underlying template and explore its variations. An extensive analysis of visual memes is performed by Zannettou and coworkers²⁹ exploiting perceptual hashing to cluster visual memes together and explore the connections between the meme content and the communities in which it circulates. In³⁰ a deep-learning classifier for memes is proposed to explore the role of memes instead of non-meme images during elections. These investigations developed relevant insights and tools for handling and researching the world of internet memes. Nevertheless, little attention is given to the fundamental aspects of the evolution of memes in terms of visual features and conveyed information. Eventually, no evidence is reported for the hypothesis of internet memes constituting a metalanguage of the internet^{10,11}.

In this work, we investigate the general evolution pattern of memes as an online communication artifact. To this aim, we leverage the evolutionist approach to define and measure the evolution rate, i.e., the number of new templates that appear online per time unit, the variation rate, i.e., the number of new instances of the same template that are produced in time; this quantity is particularly relevant concerning memes' popularity. As artistic expressions have been effectively investigated exploiting network science and physics concepts³¹, we compute the trajectory of memes in the entropy-complexity plane. Specifically, these measures, grounded on the physics of complex systems, have been employed to investigate painting arts³², revealing a temporal pattern towards higher complexity.

The basis of this investigation is a massive dataset of 2 million Reddit memes over ten years. Each image has been classified and ascribed to a template through a Machine Learning pipeline composed by an unsupervised Deep-learning based classifier followed by a density-based clustering algorithm (see Methods). Our investigation shows that the memes ecosystem size is exponentially increasing, with a doubling time of approximately six months, indicating that replication is currently the leading process. Concerning selection, we observe that memes' persistence is dominated by rapid early adoption. The variation pattern is captured by the trajectory in the entropy-complexity plane. Similarly to what happens in painting arts, we observe a tendency towards structures with increasing visual complexity; early memes were made up of simple foreground images (e.g., animals or explicit human expressions) on plain backgrounds, while later ones involve more articulated scenes (e.g., modified movie frames).

As cultural signs, memes are strictly connected to the broader cultural system in which they are embedded. While their ultimate theoretical definition is still elusive and debated in terms of methodological frames, our results indicate that memes appear as one of the most productive and adaptable areas of digital communication, functioning as a metalanguage of cultural dynamics and evolving in progressive forms of textual complexity.

Results

Our study starts from Dawkins' hypothesis of the meme as the basic unit of cultural evolution, in connection with the post-memetics analyses of memes as cultural signs. We studied the evolution of Internet visual memes, i.e. images with (typically) overlapped text strings over a time span of 10 years. The dataset comprises around 2 million images, that were grouped together exploiting an unsupervised Machine Learning routine (see Methods). Such an extended dataset enables us to investigate some properties of this particular cultural meme. The clusters retrieved by our procedure correspond to the *templates* of the various memes. In the following, we refer to "meme" as for the template, while each image belonging to a given template is an "instance" of the meme.

To quantify the growth of memes adoption we computed their evolution rate. For each cluster, we store the creation time of its first instance. Next, per each week of sampling, we compute the number of new templates. The result is reported in Fig. 1, in which the growth rate is estimated by an exponential fit, giving a doubling time for the number of templates $T \sim 6$ months. The sudden drop in the plot is a finite size effect of the dataset. Namely, for subreddits `r/memes` and `r/dankmemes` it has not been possible to download the data over 2019, due to the exponential trend in the number of memes (see Fig. 4). The fit was performed considering the data until January 2019.

Another relevant quantity in terms of cultural evolution is the *mutation rate*. The mutation rate can be approximated by looking at the instances of each meme. For each template we computed the distribution of the differences (Δt) between the creation times of an instance and the following one. This distribution reveals the nature of growth of each cluster: a distribution skewed towards low values of Δt corresponds to a very fast and bursty growth dynamic. Conversely, larger values of Δt may reveal a more persistent template, whose instances occur more spaced in time. In Fig. 2 the distribution of the "inter-instance" times (blue histograms, left column) is reported together with the clusters lifetime distribution (orange histograms, right column) i.e. the time interval between the first and the last instance of a given meme. These distributions are computed for different typical cluster sizes (indicated as CS).

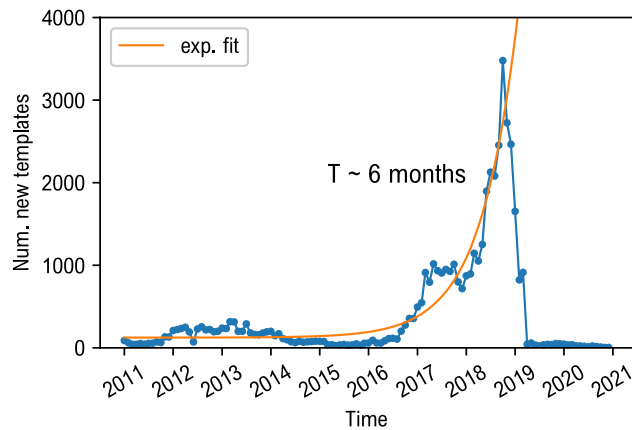


Figure 1. Evolution rate of internet memes. The number of new templates per each month is reported as a function of time. The growth rate is estimated through an exponential fit, with a doubling time of $T \sim 6$ months.

Overall, lifetime results positively correlated with cluster size. Small clusters show a heterogeneous distribution of instances' inter-arrival times, comprising both small-bursty clusters and small-slowly paced ones. The lifetime is peaked towards low values. This behaviour is also shown by very recent clusters whose actual size cannot be estimated within our dataset, as their evolution is ongoing. As the cluster size grows, we observe a shift of the inter-times distribution towards low values, unveiling faster dynamics, while the lifetime distribution is concentrated around larger values. By looking the corresponding growth curves we observe an ensemble of trajectories that tend to display a fast initial build up of popularity, followed by a slower diffusion that determines the longer lifetime values. Conversely there are also examples of memes that takes more time to reach a wide popularity. This aspect may be due to non-trivial popularity dynamics, calling for further research.

Following Sigaki et al.³², we investigated the evolution of memes in the entropy-complexity plane. For each meme instance we computed the values of $H(P)$ and $C(P)$ and then averaged the obtained values by year. The results are reported, for each subreddit, in Fig. 3. We observe that each community moves towards higher complexity values, except for *r/AdviceAnimals*, whose posting rules limit the natural evolution of produced memes; this effect could be also linked to the overall decrease in memes production observed for this community. Interestingly, also paintings followed a similar trajectory in the entropy-complexity plane: quite localized along the entropy axis, but shifting towards higher complexity in time.

The tendency of memes to evolve towards more complex structures can be explained considering this object as part of the emerging internet meta-language. In fact, memes are used to quickly vehicle context-specific content, which in turn evolves towards more and more specific templates. This may lead to a segregation effect, with a specific dialect depending on the community in which a meme is shared. In fact a meme created for a specific community, e.g. gaming community, does not have to be universally comprehensible across the web. This aspect leads to the use of more complex and specific patterns.

Discussion

The Internet provides an environment in which information quickly spreads and adapts to comply with users' cognitive abilities. A foundational question about memes is whether they present culturally and temporally transcendent characteristics in their organizing principles and how they evolve. Such a significant increase and spread of visual memes can be read under the light of post-memetics theories. Visual memes are favored by the rapid, fluid, continuously changing internet environment because of their simplicity, ease of handling and broad applicability in terms of subjects and situations. We find support for the hypothesis that memes are part of an emerging form of internet metalanguage: on one side, we observe an exponential growth with a doubling time around 6 months; on the other side, the complexity of memes contents increases, allowing to timely represent social trends and attitudes. Our analysis shows that memes are relational entities functioning as flexible elements of a metalanguage that de-codifies and re-codifies the cultural system. They appear as fundamental components of an organic process that affects and conditions the digital environment and produces evolving forms of visual and textual complexity.

Methods

Data breakdown. Reddit is an online social media platform that aggregates users in communities of interest. In the last years, it has been widely employed to perform academic research on online communities, and the number of active users on this platform is constantly increasing³³.

The visual memes used as dataset for this study were downloaded through the Pushshift Reddit Dataset³⁴, selecting four communities (subreddits) explicitly devoted to share and discuss about memes, namely: *r/AdviceAnimals*, *r/memes*, *r/CemeteryComedy* and *r/dankmemes*. Data were collected considering a ten years window, from 2011 to 2020. In Fig. 4 the number of downloaded posts per each community is reported, as function of time. Not all the communities started their activity simultaneously. A list of all images urls employed in this study is available online³⁵.

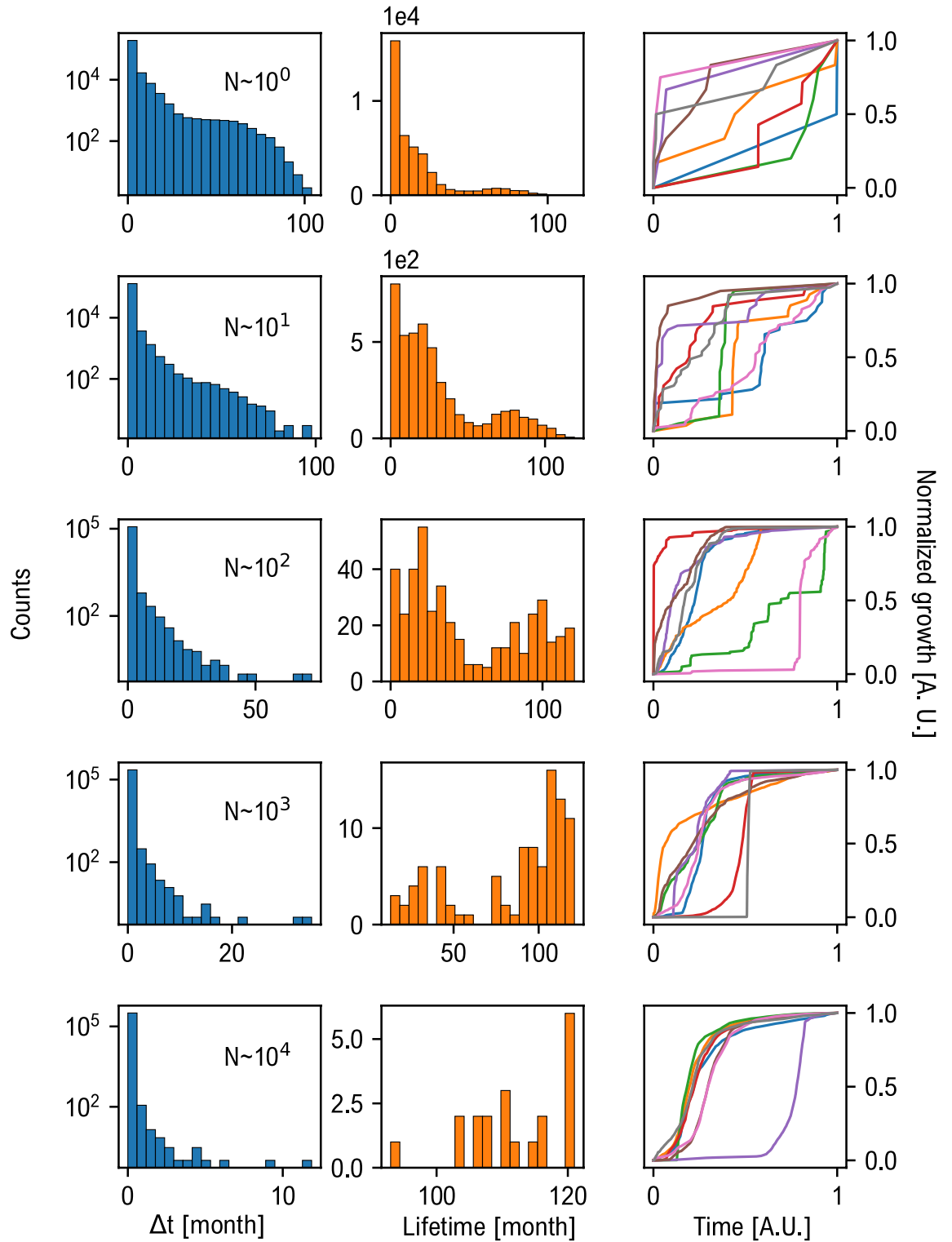


Figure 2. Mutation rate of memes. Left column: distribution of instances' inter-arrival times (Δt); central column: lifetime distribution of memes; right column: exemplary growth curve of meme adoption. Each row correspond to a typical size of meme cluster. Interestingly, the larger the cluster size the longer the lifetime. However, long lasting memes template shows skewed distributions for the inter-arrival times (Δt). This highlights the tendency of long lasting memes to undergo strong early adoption.

Clustering. One of the main features of visual memes is their recurrent nature: starting from an initial template, memes are produced through text or image modifications resulting each time in a new instance that stems

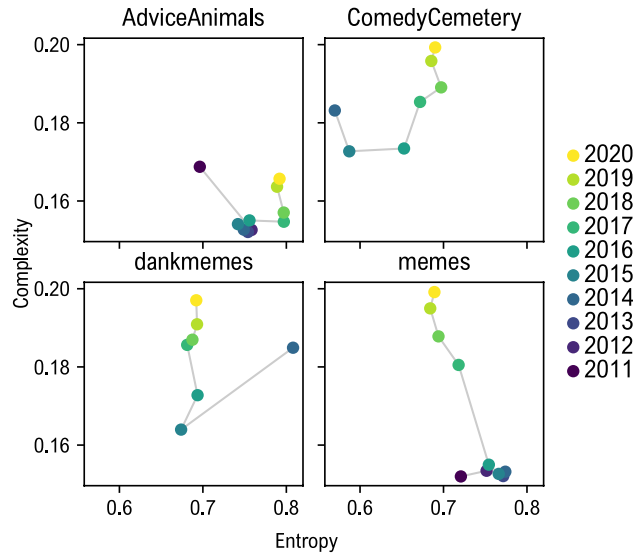


Figure 3. Trajectories in entropy-complexity plane for the four Reddit communities. All, except AdviceAnimals present an evolution towards higher values of complexity that resembles that of painting arts³². Each dot is the average value of entropy and complexity for each year. Visually complex templates could be linked to the tendency of using memes to continuously represent social trends and attitudes, supporting the hypothesis that memes are part of the emerging internet meta-language.

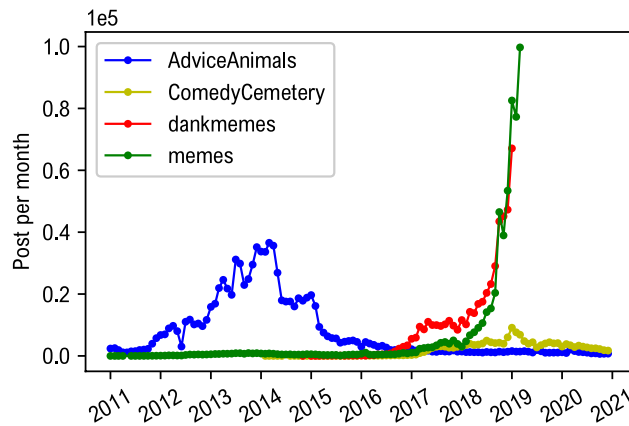


Figure 4. Dataset used in this work. Each curve shows the number of posts per month that were downloaded from Reddit. Each posts correspond to an image. The total amount of downloaded memes is about 2 million.

from the original template. To measure the evolution and variation rate of visual memes, it is crucial to cluster them according to the underlying template. As the collected number of memes is about two million, such a large amount of images calls for automatic classification methods.

Our unsupervised clustering procedure is divided into two steps: first, we apply, to our knowledge, the state of the art Deep Learning implementation for unsupervised image clustering, that is called SCAN³⁶ (Semantic Clustering by Adopting Nearest neighbors), followed by a further clustering procedure through the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm³⁷.

Specifically, the SCAN algorithm works in two steps. In the first one, self-supervised learning is used to train a neural network with parameters ϑ that maps images $(x_i, i = 1 \dots N)$ into feature representations $\phi_\vartheta(x_i) \in \mathbb{R}^d$. Therefore, each image is represented by a vector of dimension d , with d being the dimension of the embedding space carrying semantically meaningful information about its content. The parameters ϑ are determined by minimizing the loss function given by the distance δ between the representation of the image and the representation of their augmentations:

$$\min_{\vartheta} \delta(\phi_\vartheta(x_i), \phi_\vartheta(T[x_i])) .$$

The augmentation of an image may be a rotation, an affine or perspective transformation, etc.

In the second step, a neural network with parameters η is used to classify an image x_i and its nearest neighbors sampled exploiting its corresponding representation $\phi_\eta(x_i)$. As the second task is a classification, the output is a probability distribution over the considered classes for the given image: $\phi_\eta(x) \in [0, 1]^C$, where C is the number of clusters considered.

The loss function, in this case, is made of two terms

$$\Lambda = -\frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} \sum_{k \in \mathcal{N}_x} \langle \phi_\eta(x), \phi_\eta(k) \rangle + \lambda \sum_{x \in \mathcal{C}} \phi_\eta^c \log(\phi_\eta^c),$$

$$\text{with } \phi_\eta^c = -\frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} \phi_\eta^c(x),$$

where \mathcal{D} is the dataset comprising all images, \mathcal{N}_x is the set of nearest neighbors of image x , \mathcal{C} is the set of clusters, and $\langle \cdot \rangle$ is the dot product. The first term aims to maximize the probability that the image and its nearest neighbors are classified in the same class. The second term avoids the formation of a single cluster containing all the images and forces the spread of the predictions uniformly across the clusters. Eventually, for each processed image, we get a representation vector of size $d = 2048$ and a set to which it belongs.

The SCAN algorithm provides us with an informative and compact representation for each image in our dataset together with a first, high-level clustering. SCAN divides the corpus into four clusters that group the visual memes into three broad and general categories: animals (two sets), humans, and others. The two clusters with animal images have been merged together. In other words, memes containing humans represent the 50% of the corpus, followed by 25% of animals and 25% of other kind of contents.

To obtain a template-based clustering, we exploited HDBSCAN^{37,38}. For each high-level cluster obtained from SCAN, the entire corpus of memes can be represented by a matrix whose rows are the representations $\phi_\eta(x_i)$. To make the problem computationally more tractable, Principal Component Analysis was used to reduce the features' dimension from 2048 to 20. This allowed us to better fit our computational resources and did not cause any reduction in the quality of the clustering. By applying HDBSCAN to such a matrix, we were able to get a label for each image of our corpus and to separate the memes by their template. Notably, HDBSCAN can separate clusters from noisy points. Part of the corpus does not belong to any template, and therefore is marked as noise and grouped in a large "cluster of noise". Despite this suitable property of the algorithm, some clusters may result made up of images whose template is not the same for all. A purity measurement is therefore required to exclude from the analysis too heterogeneous clusters, i.e. clusters below a given purity threshold.

In a completely unsupervised framework, the quality of clustering is in general not easy to evaluate³⁹. A quantity that is usually employed as an objective function for clustering is the average-pairwise distance \bar{S}_k , which evaluates the intra-cluster homogeneity⁴⁰, i.e. how much each element of the cluster is, on average, similar to all the others. Its definition is given by

$$\bar{S}_k = \frac{1}{N_k^2} \sum_{x_i, x_j \in \mathcal{C}_k} \|\phi_\eta(x_i) - \phi_\eta(x_j)\|^2.$$

In our case, we employed the latter quantity to measure the purity of the clusters identified by HDBSCAN, together with the cluster size. In Fig. 5(a) for each cluster is reported \bar{S}_k and the cluster size. The red dots are the clusters that are not considered for the following analysis. Four of them correspond to the "noisy clusters" retrieved by HDBSCAN. All of them result in outliers with respect to the joint size and \bar{S}_k distribution, whose marginals distributions are reported in panels (b) and (c) respectively. HDBSCAN performances depends upon several parameters. We used mainly default values, setting `min_cluster_size` and `min_samples` to 20 and 3, respectively.

Entropy and complexity. Permutation entropy H and statistical complexity C are two quantities that can be used to synthesize general properties of images, based on the value and relative disposition of their pixels. In the following, we give a minimal description of both quantities, and we refer the reader to original articles for more formal details^{41–45}. Permutational entropy measures the degree of disorder in the pixel arrangement. High values indicate high pixel randomness, while low values correspond to more regular patterns. Statistical complexity instead measures the amount of "structural" complexity. Non-trivial spatial patterns give rise to positive values, while extremely ordered or disordered patterns correspond to low values.

To compute H and C all colored images were converted to grayscale. Thus, each image consists of two-dimensional matrix. Next, for all the 2×2 submatrices comprised in the image the relative ordering of the pixel is computed³². For a collection of four elements, a number of $4! = 24$ possible permutations can be obtained. By counting the relative occurrence of each permutation, a probability mass function

$$P = \{p_1 \dots p_N\} \quad \text{with} \quad \sum_{i=1}^N p_i = 1,$$

can be built. Shannon's entropy is then computed over this probability distribution to obtain the permutation entropy

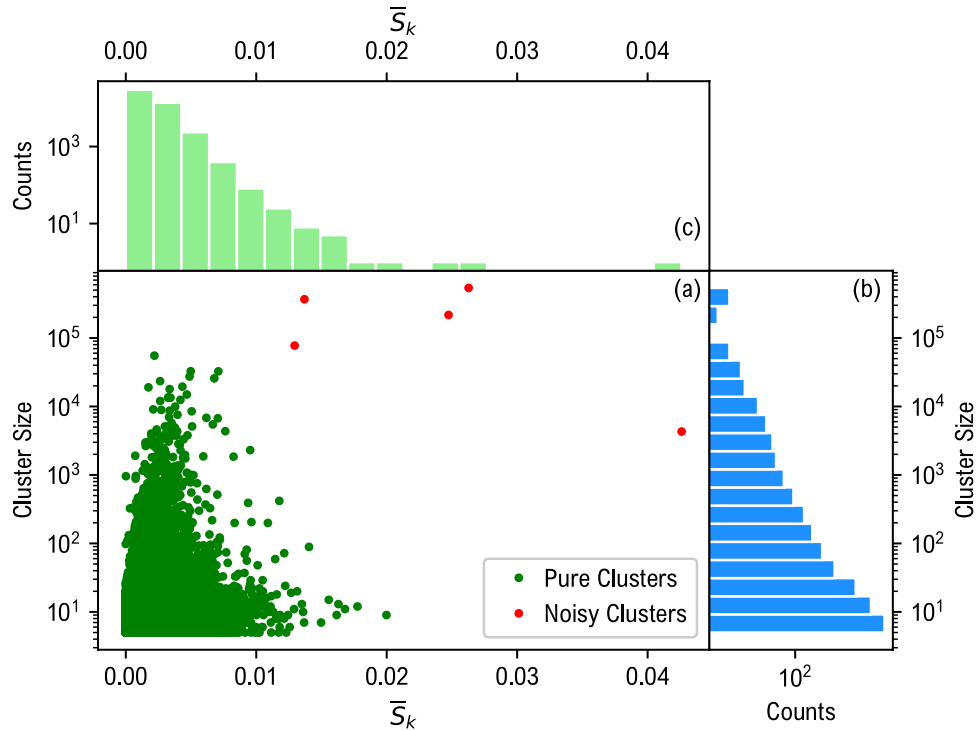


Figure 5. Panel (a): joint distribution of average-pairwise distance \bar{S}_k and cluster size. Each dot corresponds to a cluster, i.e. a template for memes. Red dots represent “noisy clusters” identified by HDBSCAN that are outliers with respect to the cluster size and \bar{S}_k distributions, shown respectively in panels (b) and (c). Noisy clusters are removed from the analysis.

$$H(P) = \frac{S(P)}{\log(N)} = \frac{1}{\log(N)} \sum_{i=1}^N p_i \log\left(\frac{1}{p_i}\right). \tag{1}$$

Given the probability mass function P , its discrepancy with respect to a uniform distribution U

$$U = \{u_1 \dots u_N\} \quad \text{with} \quad \sum_{i=1}^N u_i = 1,$$

is obtained by computing the Jensen-Shannon divergence

$$D(P, U) = S\left(\frac{P + U}{2}\right) - \frac{S(P)}{2} - \frac{S(U)}{2}.$$

Combining this quantity with $H(P)$, the statistical complexity can be computed as

$$C(P) = \frac{D(P, U) H(P)}{D^*}, \tag{2}$$

with the normalizing factor

$$D^* = \max_P D(P, U) = -\frac{1}{2} \left[\frac{N+1}{N} \log(N+1) + \log(N) - 2 \log(2N) \right].$$

The evolution of visual patterns can be studied as a trajectory in the above defined entropy-complexity plane^{42,43}, following the same approach used for paintings³².

Data availability

The list of all the images used in this study is available online³⁵. Software used for computations are all properly referenced in the main text.

Received: 2 July 2021; Accepted: 22 September 2021
 Published online: 08 October 2021

References

- Vicario, M. D. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**, 554–559 (2016).
- Schmidt, A. L., Zollo, F., Scala, A., Betsch, C. & Quattrociochi, W. Polarization of the vaccination debate on facebook. *Vaccine* **36**, 3606–3612 (2018).
- Sunstein, C. R. *#republic* (Princeton University Press, New York, 2017).
- Del Vicario, M. *et al.* Echo chambers: Emotional contagion and group polarization on facebook. *Sci. Rep.* **6**, 1–12 (2016).
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociochi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci.* **118**, e2023301118 (2021).
- Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**, 9216–9221 (2018).
- Dawkins, R. *28. The Selfish Gene* 140–142 (Princeton University Press, New York, 2014).
- Deacon, T. W. Editorial: Memes as signs: The trouble with memes (and what to do about it). *Semiot. Rev. Books* **10**, 1–3 (1999).
- Sebeok, T. A. & Danesi, M. *The Forms of Meaning* (De Gruyter, Cambridge, 2000).
- Cannizzaro, S. Internet memes as internet signs: a semiotic view of digital culture. *Sign Syst. Stud.* **44**, 562–586 (2016).
- Fomin, I. Memes, genes, and signs: Semiotics in the conceptual interface of evolutionary biology and memetics. *Semiotica* **2019**, 327–340 (2019).
- Distin, K. *The Selfish Meme* (Cambridge University Press, Cambridge, 2004).
- Leskovec, J., Backstrom, L. & Kleinberg, J. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, 497–506 (ACM Press, 2009).
- Ienco, D., Bonchi, F. & Castillo, C. The meme ranking problem: Maximizing microblogging virality. In *2010 IEEE International Conference on Data Mining Workshops*, 328–335 (IEEE, 2010).
- Romero, D. M., Meeder, B. & Kleinberg, J. Differences in the mechanics of information diffusion across topics. In *Proceedings of the 20th international conference on World wide web - WWW '11*, 695–704 (ACM Press, 2011).
- Bauchhage, C. Insights into internet memes. In *Proceedings of the International AAAI Conference on Web and Social Media* Vol. 5 42–49 (2011).
- Ratkiewicz, J. *et al.* Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web - WWW '11*, 249–252 (ACM Press, 2011).
- Weng, L., Flammini, A., Vespignani, A. & Menczer, F. Competition among memes in a world with limited attention. *Sci. Rep.* **2**, 335 (2012).
- Coscia, M. Competition and success in the meme pool: a case study on quickmeme.com. <https://arxiv.org/abs/1304.1712> (2013).
- Coscia, M. Average is boring: How similarity kills a meme's success. *Sci. Rep.* **4**, 6477 (2014).
- Coscia, M. Popularity spikes hurt future chances for viral propagation of protomemes. *Commun. ACM* **61**, 70–77 (2017).
- Weng, L., Menczer, F. & Ahn, Y.-Y. Predicting successful memes using network and community structure. In *Proceedings of the International AAAI Conference on Web and Social Media* Vol. 8 535–544 (2014).
- Ferrara, E. *et al.* Clustering memes in social media. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, 548–555 (2013).
- Dang, A., Moh'd, A., Gruz, A., Milios, E. & Minghim, R. A visual framework for clustering memes in social media. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 713–720 (2015).
- Tsur, O. & Rappoport, A. Don't let me be# misunderstood: Linguistically motivated algorithm for predicting the popularity of textual memes. In *Proceedings of the International AAAI Conference on Web and Social Media* Vol. 9 426–435 (2015).
- Adamic, L. A., Lento, T. M., Adar, E. & Ng, P. C. Information evolution in social networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM '16*, 473–482 (Association for Computing Machinery, New York, NY, USA, 2016).
- Kuhn, T., Perc, M. & Helbing, D. Inheritance patterns in citation networks reveal scientific memes. *Phys. Rev. X* **4**, 041036 (2014).
- Dubey, A., Moro, E., Cebrian, M. & Rahwan, I. Memesequencer: Sparse matching for embedding image macros. In *Proceedings of the 2018 World Wide Web Conference, WWW '18*, 1225–1235 (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018).
- Zannettou, S. *et al.* On the origins of memes by means of fringe web communities. In *Proceedings of the Internet Measurement Conference 2018, IMC '18*, 188–202 (Association for Computing Machinery, New York, NY, USA, 2018).
- Beskow, D. M., Kumar, S. & Carley, K. M. The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. *Inf. Process. Manag.* **57**, 102170 (2020).
- Perc, M. Beauty in artistic expressions through the eyes of networks and physics. *J. R. Soc. Interface* **17**, 20190686 (2020).
- Sigaki, H. Y. D., Perc, M. & Ribeiro, H. V. History of art paintings through the lens of entropy and complexity. *Proc. Natl. Acad. Sci.* **115**, E8585–E8594 (2018).
- Medvedev, A. N., Lambiotte, R. & Delvenne, J.-C. The anatomy of reddit: An overview of academic research. In *Dynamics On and Of Complex Networks III*, 183–204 (Springer International Publishing, 2019).
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The pushshift reddit dataset. *Proc. Int. AAAI Conf. Web Social Media* **14**, 830–839 (2020).
- Valensise, C. M. Cdcs repository. <https://github.com/cdcslab/MemesEvolution>.
- Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M. & Van Gool, L. Scan: Learning to classify images without labels. In *Proceedings of the European Conference on Computer Vision*, 123550273 (2020).
- Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*, 160–172 (Springer Berlin Heidelberg, 2013).
- McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Sour. Softw.* **2**, 205 (2017).
- Mehta, P. *et al.* A high-bias, low-variance introduction to machine learning for physicists. *Phys. Rep.* **810**, 1–124 (2019).
- Rokach, L. & Maimon, O. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*, 321–352 (Springer-Verlag).
- Bandt, C. & Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **88**, 174102 (2002).
- López-Ruiz, R., Mancini, H. & Calbet, X. A statistical measure of complexity. *Phys. Lett. A* **209**, 321–326 (1995).
- Ribeiro, H. V., Zunino, L., Lenzi, E. K., Santoro, P. A. & Mendes, R. S. Complexity-entropy causality plane as a complexity measure for two-dimensional patterns. *PLoS ONE* **7**, e40689 (2012).
- Zunino, L. & Ribeiro, H. V. Discriminating image textures with the multiscale two-dimensional complexity-entropy causality plane. *Chaos, Solitons & Fractals* **91**, 679–688 (2016).
- Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A. & Fuentes, M. A. Distinguishing noise from chaos. *Phys. Rev. Lett.* **99**, 154102 (2007).

Author contributions

C.M.V. performed research; C.M.V., A.S., M.C. and W.Q. designed research; all authors participated to internal revision and wrote the paper.

Competing interest

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.M.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021