

ROYAL SOCIETY OPEN SCIENCE

 Open Access

 Check for updates

Research articles

How networks shape diversity for better or worse

Andrea Musso  and Dirk Helbing

Published: 22 May 2024 | <https://doi.org/10.1098/rsos.230505>

 Review history

Review history is available via Web of Science at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1098/rsos.230505>

Abstract

Socio-diversity, the variety of human opinions, ideas, behaviours and styles, has profound implications for social systems. While it fuels innovation, productivity and collective intelligence, it can also complicate communication and erode trust. So what mechanisms can influence it? This paper studies how fundamental characteristics of social networks can support or hinder socio-diversity. It employs models of cultural evolution, mathematical analysis and numerical simulations. We find that pronounced inequalities in the distribution of connections obstruct socio-diversity. By contrast, the prevalence of close-knit communities, a scarcity of long-range connections, and a significant tie density tend to promote it. These results open new perspectives for understanding how to change social networks to sustain more socio-diversity and, thereby, societal innovation, collective intelligence and productivity.

1. Introduction

In his seminal work, *The Selfish Gene* [1], Richard Dawkins proposes a compelling perspective on culture and its evolution. He argues that, by viewing *memes*¹—a collective term encompassing human ideas, opinions, behaviours and styles—as cultural counterparts to genes, one can use the foundational principles of biological evolution to explain the origins and development of cultural traits.

Biological evolution is driven by mutation and selection, acting on biology's fundamental units: genes [2]. Mutation introduces new genetic variants into a population, which undergo selection through individual interactions. The nature of these interactions varies: some individuals engage with each other frequently, while others are predominantly isolated. Such patterns of selective interaction, commonly known as *networks*, are fundamental to understanding the evolution of complex systems. [Back to top](#)

interaction network structure, critically determine survival chances, thereby influencing which genetic variants are perpetuated through reproduction and which become extinct. In essence, the structure of the interaction network is instrumental in shaping the variety of genes present in the population, or, in other words, its bio-diversity.

Research suggests a complex relationship between the structure of interaction networks and bio-diversity. For example, certain inter-individual interaction networks can increase the advantage of fitter individuals, potentially reducing bio-diversity [2,3]. By contrast, the pronounced nestedness found in inter-species interaction networks seems to boost bio-diversity by reducing direct competition among species [4].

Socio-diversity, defined as the variety of memes (i.e. ideas, opinions, behaviours and styles) present in a society, is a social parallel to bio-diversity. Dawkins' perspective on cultural evolution suggests that socio-diversity emerges from the interplay of imitation and innovation, acting upon culture's basic units: memes [5,6]. Innovation, akin to biological mutation, creates new cultural variants (new memes). Imitation, as a counterpart of biological selection, determines which variants diffuse across society.

Just as the structure of species interaction networks influences bio-diversity, the structure of social interaction networks affects socio-diversity. There is substantial evidence that the network structure in which individuals are embedded [7] significantly influences the diffusion of various memes, such as obesity [8], smoking [9], cooperation [2,10,11] and product adoption [12]. Generally, clustered networks, characterized by numerous closed triangles (i.e. your friends are also friends), excel at spreading memes, such as complex behaviour, requiring repeated endorsement [12–14]. The dense local connections in these networks provide the repeated exposure necessary for such memes to take hold. Conversely, networks characterized by an abundance of long-range connections are more effective at disseminating simpler memes that need minimal reinforcement [15,16]. These far-reaching connections enable quick meme transmission across diverse network areas, facilitating rapid spread.

Network structure also significantly shapes meme creation. Efficient networks, characterized by short distances between nodes, appear to hinder the creation of radically novel memes by facilitating blind imitation [17,18]. By contrast, networks with larger average distances between nodes, seem to foster the generation of novelty by providing fewer imitation opportunities [17,18].

In summary, while there is substantial evidence on the influence of network structure on patterns of meme diffusion and creation, and these patterns clearly affect socio-diversity, research directly examining the relationship between network structure and socio-diversity is scarce. As a result, a systematic understanding of this relationship is still lacking, especially when compared with the comprehension in ecology and conservation biology of the similar relationship between network structure and bio-diversity.² Crucial questions remain open: How can we measure a social network's potential to foster socio-diversity? Which social networks enhance socio-diversity, and which ones diminish it?

This paper addresses these important questions by introducing a novel index linking network structure and socio-diversity: the *structural diversity index*. This index, derived from the theory of random walks on networks [19,20], quantifies the propensity of a network to support socio-diversity; its ability to protect unpopular memes from being crushed by more popular ones. With our novel index, understanding if a social network enhances or diminishes socio-diversity becomes straightforward: it suffices to compare the network's index value against a benchmark (such as the complete network). If a network's index value is higher than the benchmark, then the network promotes socio-diversity; otherwise, it hinders it.

The index is designed for scalability, can handle large networks and is easily accessible through the Python package accompanying this paper.

Employing our novel index, we conducted an extensive exploration of the relationship between network structure and socio-diversity across a broad range of real-world and synthetic networks. We focused on some key network characteristics, such as the shape of the degree distribution, the edge density and the prevalence of long-range connections. Our selection of these characteristics is underpinned by four guiding hypotheses. First, networks dominated by highly connected individuals may see diminished socio-diversity as a result of the disproportionate influence on cultural spreading exerted by such individuals. Second, networks with prevalent long-range connections might display lower socio-diversity due to the homogenizing effects of such ties. The pervasive spread of global pop culture and its consequential effect on local cultures exemplifies the possible homogenizing impact of these long-range connections [21]. Third, networks with a high density of connections may bolster socio-diversity. Specifically, increasing the number of connections diminishes the influence of each individual connection and, thereby, lowers the chances of viral meme cascades [12]. Fourth, networks characterized by numerous close-knit communities may enhance socio-diversity. In fact, when these communities share homogeneous memes, they limit exposure to new memes and foster the persistence of those already adopted.

We initiated our investigation by examining two prominent synthetic network models: scale-free [22] and Watts–Strogatz [23] networks. Although these models might seem simplistic, they serve as useful tools to isolate the effect of different network characteristics. Specifically, scale-free networks offer a lens to study degree-heterogeneity, or in simpler terms, the unequal distribution of connections among nodes. Our analysis of these networks' structural diversity index suggests that such a high disparity in connections can reduce socio-diversity. This aligns with the hypothesis that individuals with numerous connections may inadvertently suppress socio-diversity due to their pronounced influence on cultural transmission. Conversely, Watts–Strogatz networks provide a framework to understand the ramifications of long-range connections and close-knit communities. Our findings indicate that fewer long-range connections and more close-knit communities are positive for socio-diversity. This is compatible with the idea that cultural convergence towards a 'global village' is expected to erode socio-diversity.

Moving beyond these simplistic models, we broadened our investigation to include hundreds of real-world networks. Using the comprehensive real-world network database provided by graph-tool [24], we computed the structural diversity index of networks originating from various social contexts. Our findings reinforce that high inequality in the distribution of connections suppresses socio-diversity, while the scarcity of long-range connections and a prevalence of close-knit communities amplify it. Furthermore, we observed that a high density of connections also tends to improve socio-diversity.

Although the factors shaping socio-diversity have drawn interest from social scientists [25–29], its role in social science research has not reached the prominence of bio-diversity in ecology and conservation biology. Yet, socio-diversity has profound real-world implications [30]. On the positive side, socio-diversity is a catalyst for innovation [31], promotes cooperation [32,33], and can increase productivity [34,35]. Research shows that protecting novel and rare ideas from premature dismissal [17,36] and encouraging independent thought [37–40] can enhance a group's ability to solve all kinds of problems, making it more 'intelligent', more productive and more innovative. Conversely, socio-diversity can pose challenges to group cohesion [41,42], impede effective communication [43], and erode trust [44,45]. Our research seeks to reduce the relative disparity in attention between bio- and socio-diversity by shed light on the interplay between network structure and socio-diversity.

2. Results

2.1. Structural diversity index

PDF

Help



Back to top

Consider a social network represented abstractly by a connected undirected graph G . In this representation, each vertex stands for an individual, and edges symbolize undirected and mutual relationships, such as friendships, acquaintances or interactions.

For illustration, imagine that two individuals, Alice and Bob, decide to play the ‘random social exploration game’, a variation of Milgram’s celebrated small-world experiment [46]. In this game, Alice and Bob each randomly select a friend from their network and send them a letter. This letter carries a simple instruction: ‘Please choose a friend at random and forward this letter to them.’ Every recipient follows this directive, passing the letter onward within their network. As the letters get forwarded again and again, they randomly explore the social circles of both Alice and Bob. The game ends when the two letters *meet*, i.e. when they simultaneously end up in the mailbox of the same individual.

As an example, consider the hexagonal-shaped social network depicted in [figure 1A](#). This network comprises six individuals: Alice (a), Bob (b), Carla (c), Darcy (d), Elon (e) and Frank (f). In the initial phase of the random social exploration game (shown in A), Alice selects a friend at random to send her letter. Thus, her letter stands an equal chance of landing with Bob, Carla, Darcy or Frank. In this particular illustration, fate dictates the letter to be sent to Frank, as shown in B. Upon receiving the letter, Frank, too, makes a random choice, deciding to forward the letter to Elon. Simultaneously, Bob’s letter also finds its way to Elon, having first been relayed through Carla. It is at this point, in Elon’s mailbox, that the letters from Alice and Bob *meet*, marking the game’s end.

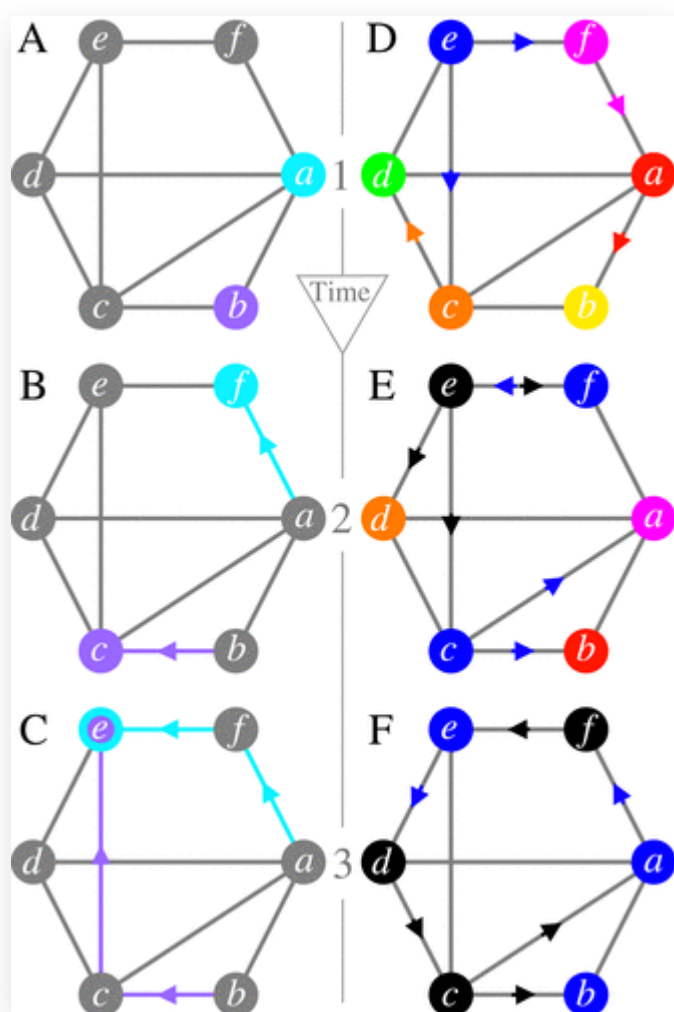


Figure 1. An illustration of the random social exploration game (A–C) and the progression of cultural evolution in a small social network (D–F). The network features six individuals: Alice (a), Bob (b), Carla (c), Darcy (d), Elon (e) and Frank (f). (A–C) In the random social exploration game, letters sent by Alice and Bob are randomly forwarded through the network until they meet in a mailbox (see main text for details). In this illustration, the letters’ journeys are marked by highlighted edges, converging at Elon’s mailbox. (D–F) These panels illustrate cultural evolution within our social network in three-time steps. Each panel depicts both the present meme distribution and its imminent evolution by arrows. An individual’s current meme is reflected by the colour of their vertex, whereas upcoming changes are represented by the colour of the arrows pointing towards their vertex. For instance, Alice’s red meme in d transforms into Frank’s magenta meme in e .

indicated by the magenta arrow pointing from Frank to Alice. It is noteworthy that in d , no arrows point towards Elon, suggesting he does not mimic others at $t = 1$, but instead introduces a new meme (the black meme).

[Download figure](#) | [Open in new tab](#) | [Download PowerPoint](#)

The *expected meeting time* of the network G , denoted $\langle M_G \rangle$, is defined as the average number of forwards required for the letters to meet. Specifically, it is computed by conducting many repetitions of the random social exploration game, with letters starting out from different individuals and then averaging the number of forwards necessary for the letter to meet in each game iteration. Formally, the expected meeting time $\langle M_G \rangle$ is defined as the average number of steps before two uniformly started random walks on G visit the same vertex simultaneously. It is a well-studied network statistic [47,48].

The structural diversity index of the network G is defined as the ratio between its expected meeting time $\langle M_G \rangle$ and its number of vertices, or size, $|V(G)|$:

$$\Delta(G) = \frac{\langle M_G \rangle}{|V(G)|}. \tag{2.1}$$

On the surface, $\Delta(G)$ is a scale invariant measure of the ease of meeting during a random walk on the network. Although this metric is solely based on the network’s structure, it offers powerful predictions of the network’s propensity to support socio-diversity.

The fascinating relationship between the structural diversity index $\Delta(G)$ and socio-diversity is best understood through a simple model of cultural evolution, commonly known as the voter model [49]. As before, we have a population occupying the vertices of a network G with edges symbolizing various undirected and reciprocal relationships, such as friendships, acquaintances or interactions.

At the beginning, each individual i of the population displays a distinct meme m_i . A meme symbolizes an individual cultural trait that can vary—such as political beliefs or musical tastes. Cultural evolution is assumed to happen at discrete time steps. During each step, individuals simultaneously modify their meme in one of two ways:

- (i) by imitating the current meme of a randomly chosen neighbour, which occurs with a probability of $1 - r$.
- (ii) by inventing a novel meme that has not been previously existed, with a probability of r .³

The parameter r , with $0 \leq r \leq 1$, is referred to as the innovation rate. It measures the equilibrium between the two key evolutionary forces of imitation and innovation. Higher values of r stimulate innovation, while lower values strengthen imitation.

Figure 1D–F illustrates cultural evolution in a small social network over three time-steps $t = 1, 2, 3$. In d , each individual displays a distinct meme, represented by the colour of its respective vertex. We focus on the journey of a specific individual, Elon. Transitioning from $t = 1$ (d) to $t = 2$ (e), we observe that Frank and Carl imitate Elon’s blue meme (indicated by blue arrows). By contrast, Elon innovates, introducing a new meme: the black meme. Globally, the meme landscape has undergone a transformation. Elon’s black meme enters the scene, the green and yellow disappear, and the blue meme gets increased attention. Transitioning from $t = 2$ (e) to $t = 3$ (f), Elon opts to imitate Frank by embracing the blue meme. Co

Frank, Carl and Darcy find Elon's new meme appealing and imitate it. As a result, only two dominant memes emerge: blue and black.

In short, as cultural evolution unfolds over time, individuals either innovate (as Elon) by creating new memes or imitate (as everyone else) by adopting existing memes from their peers. These processes continually reshape the meme landscape in the population, influencing its overall socio-diversity. For instance, it is evident that the population in [figure 1](#)'s D is inherently more diverse than that in [f](#). Yet, to quantify this difference in diversity, one requires a specific measurement of socio-diversity.

In this study, we adopt a well-known diversity measure in ecology: Simpson's diversity index [\[50\]](#). This index takes into account both the number of memes, as well as their relative abundance. It is defined as the probability that two randomly selected individuals in the population display different memes. The value ranges from 0 to 1. When the population's memes are homogeneous, the probability that two randomly selected individuals exhibit different memes is small. Consequently, Simpson's diversity index is close to 0. Conversely, when the population's memes are diverse, the probability that two randomly selected individuals display different memes is large. Accordingly, Simpson's diversity index is close to 1.

In the context of our cultural evolution model, Simpson's diversity index at fixed time t can be computed as

$$D(t) = 1 - \sum_m p_m(t)^2. \quad 2.2$$

Herein, the summation is over all memes present at time t , and $p_m(t)$ is the fraction of individuals who display meme m at that time. For illustration, the socio-diversity $D(t=1)$ of the (highly diverse) population portrayed in [figure 1](#)D is $D(1) = 0.833$, while that of the (more homogeneous) population in [figure 1](#)F is $D(3) = 0.5$. In summary, higher values of $D(t)$ reflect greater socio-diversity within the population.

Returning to our fundamental research question, we can now frame it more precisely: how does the network structure G of a population influence its socio-diversity $D(t)$? Our answer is the following simple and elegant mathematical equation, which is a generalization of results by Aldous and collaborators [\[49,51\]](#). This equation links the meeting time in the random social exploration game (depicted in [figure 1](#)A–C) to the long-term socio-diversity in our cultural evolution model (as shown in [figure 1](#)D–F). The derivation of this equation arguably represents the most intricate part of our analysis and can be found in Methods

$$D_\infty \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s \leq t} D(s) \approx 1 - e^{-2\alpha\Delta(G)}. \quad 2.3$$

The left-hand side of the equation introduces D_∞ , the average population wide socio-diversity over an extended period of time (henceforth the population's *expected socio-diversity*). This quantity is the focal point of our exploration. We aim to understand what elements of a population's network structure affect its expected socio-diversity D_∞ .

Equation [\(2.3\)](#) reveals that the expected socio-diversity is determined by two fundamental factors: first, the per capital innovation rate $\alpha = 1/V(G)$. Unsurprisingly, a higher *per capita* innovation rate leads to greater socio-diversity; second, and most importantly, the network structure G , as captured by the structural diversity index $\Delta(G)$. When $\Delta(G)$ is high, the expected socio-diversity tends to be high as well, nearing its maximum of 1. Conversely, when $\Delta(G)$ is low, the expected socio-diversity is low, actually close to the minimum value of 0. This codependence relationship suggests that the structural diversity index

$\Delta(G)$ captures the connection between network structure, upon which it depends, and socio-diversity, which it influences.

Figure 2 plots the relationship between expected socio-diversity D_∞ and the structural diversity index $\Delta(G)$ for various real-world social networks G . In a log-log plot, a clear saturating relationship is found, which aligns with our analytical prediction based on equation (2.3) (see red line).

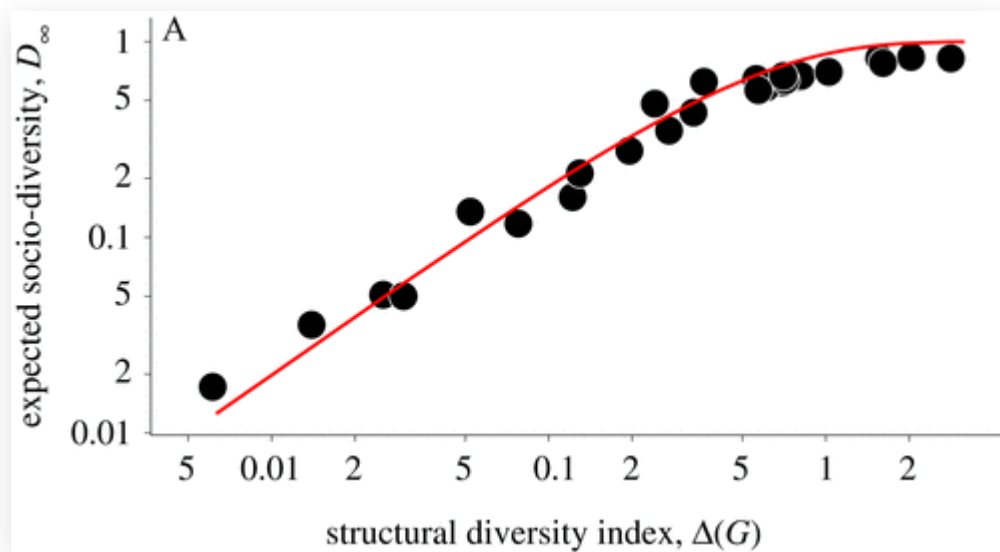


Figure 2. Relationship between expected socio-diversity D_∞ with the structural diversity index $\Delta(G)$. We simulated the cultural evolution model across various real-world social networks G to determine D_∞ . Each dot represents a simulation for a distinct social network. The red line depicts the curve $1 - e^{-2\Delta(G)}$, our theoretical estimate for D_∞ derived from equation (2.3) using $\alpha = 1$ (or $r = 1/|V(G)|$). Remarkably, the structural diversity index *predicts* expected socio-diversity levels quite accurately, as evidenced by observations scattering around the red line. See Methods for simulation parameters and descriptions of the social networks.

[Download figure](#) | [Open in new tab](#) | [Download PowerPoint](#)

2.2. Amplifiers and suppressors of socio-diversity

Equation (2.3) captures the complex interplay between socio-diversity and network structure by expressing the expected socio-diversity D_∞ as a function of a quantity that only depends on network structure, namely, the structural diversity index $\Delta(G)$. By increasing the structural diversity index $\Delta(G)$ in equation (2.3), we observe an increase in expected socio-diversity D_∞ . Hence, networks with large structural diversity index tend to favour socio-diversity, whereas those with a small one tend to obstruct it.

But what should be considered ‘large’ or ‘small’ structural diversity indices? Large and small are typically defined with respect to a benchmark. The natural benchmark here is the complete network K . The complete network reflects the total absence of social structure. There are no communities, no clusters and no differences between individual social positions; the population is structurally homogeneous. Comparing the structural diversity index of an arbitrary network G with that of an equally sized complete network K informs us about how the network structure affects the index, and, consequently, socio-diversity.

The structural diversity index of the complete network satisfies $\Delta(K) = 1$, independent of its size (see Methods for an explanation). If, for a network structure G , we have $\Delta(G) < \Delta(K) = 1$, equation (2.3) suggests that the population’s expected socio-diversity is lower than if the population were unstructured. In other words, all else being equal, the variety of memes in a population with structure G is expected to be lower than in a population with no structure. Hence, networks G with $\Delta(G) < 1$ can be said to (structurally) *suppress socio-diversity* (see [3]). By contrast, networks G with $\Delta(G) > \Delta(K) = 1$ can be said to (structurally) *amplify socio-diversity*. Indeed, according to equation (2.3), the population’s expected socio-diversity is higher than if the population were unstructured.

[Back to top](#)

Scale-free networks G_γ are networks characterized by a power-law degree distribution $P(k) \sim k^{-\gamma}$. In the Methods, we show that the structural diversity index of scale-free networks with exponent γ , with $2 \leq \gamma \leq 3$, satisfies

$$\Delta(G_\gamma) \leq |V(G_\gamma)|^{-((3-\gamma)/(\gamma-1))} \leq 1. \quad 2.4$$

Therefore, scale-free networks tend to suppress socio-diversity. Moreover, as illustrated in [figure 3a](#), diversity suppression intensifies as the scale-free network becomes more degree-heterogeneous (i.e. as the exponent γ decreases).

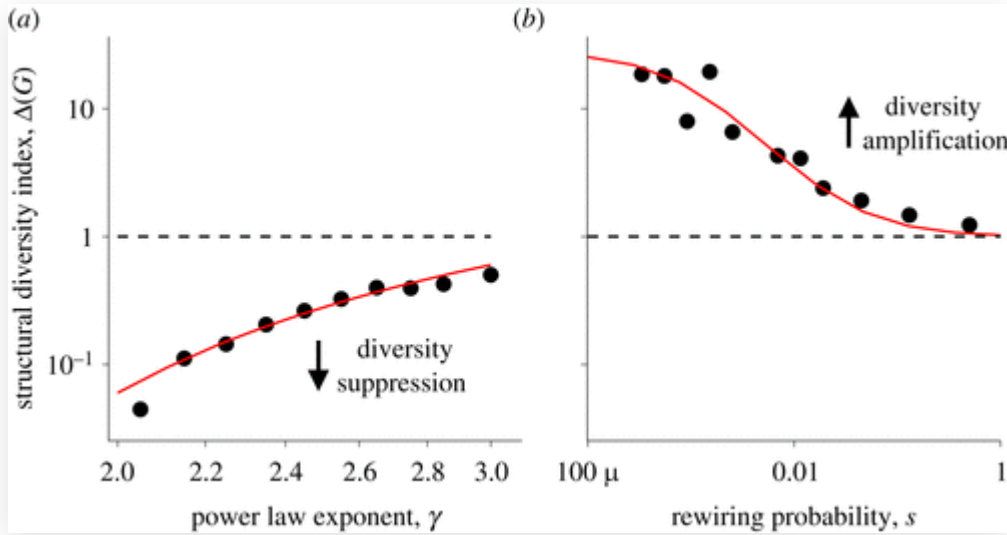


Figure 3. Numerical simulation (dots) and analytical estimates (red lines) of the structural diversity index of (a) scale-free and (b) Watts–Strogatz networks. These are plotted against (a) the power-law exponent γ and (b) the rewiring probability s . In (a), the red lines show the equation $\Delta(G_\gamma) = b \cdot |V(G_\gamma)|^{-a(3-\gamma)/(\gamma-1)}$ where a and b are obtained by ordinary least-squares fit. In (b), the red line represents the approximation in equation (2.5). Scale-free networks tend to suppress socio-diversity ($\Delta(G_\gamma) < 1$). Specifically, greater heterogeneity in the degree distribution (i.e. a smaller exponent γ) induces greater suppression of socio-diversity (i.e. smaller values of $\Delta(G_\gamma)$). Conversely, Watts–Strogatz networks tend to amplify socio-diversity ($\Delta(W_s) > 1$). However, socio-diversity amplification is reduced as more long-range connections are established or/and more randomness is inserted (i.e. as the rewiring probability s increases). Electronic supplementary material, movie S1 offers a visual comparison of cultural evolution on scale-free and Watts–Strogatz networks (see Section F.1 of the electronic supplementary material for the movie’s caption). See Methods for simulation parameters.

[Download figure](#) | [Open in new tab](#) | [Download PowerPoint](#)

Watts–Strogatz networks W_s interpolate between regular lattices and random networks by means of a parameter $s \in [0, 1]$, called rewiring probability [23]. The structural diversity index of these networks is roughly

$$\Delta(W_s) \approx \frac{s + 1/\langle k \rangle^2}{s + 1/|V(W_s)|} \geq 1, \quad 2.5$$

where $\langle k \rangle$ denotes the network’s average degree (see Methods). Hence, Watts–Strogatz networks tend to amplify socio-diversity, and this amplification weakens as randomness increases (i.e. as the rewiring probability s becomes larger)—see [figure 3b](#).

2.2.1. Characteristics of real-world networks that amplify and suppress socio-diversity

Let us broaden the scope of our analysis from the previous examples and ask: what general characteristics of networks amplify or suppress expected socio-diversity? [Figure 4a–e](#) plots the structural diversity index $\Delta(G)$ against five well-known properties of social networks—degree-heterogeneity, Wiener index, edge density, clustering and size—for a wide range of real-world [Back to top](#)

networks G . [Table 1](#) presents the outcomes of five regression models. These models help quantify the correlations shown in [figure 4](#) and evaluate their robustness.

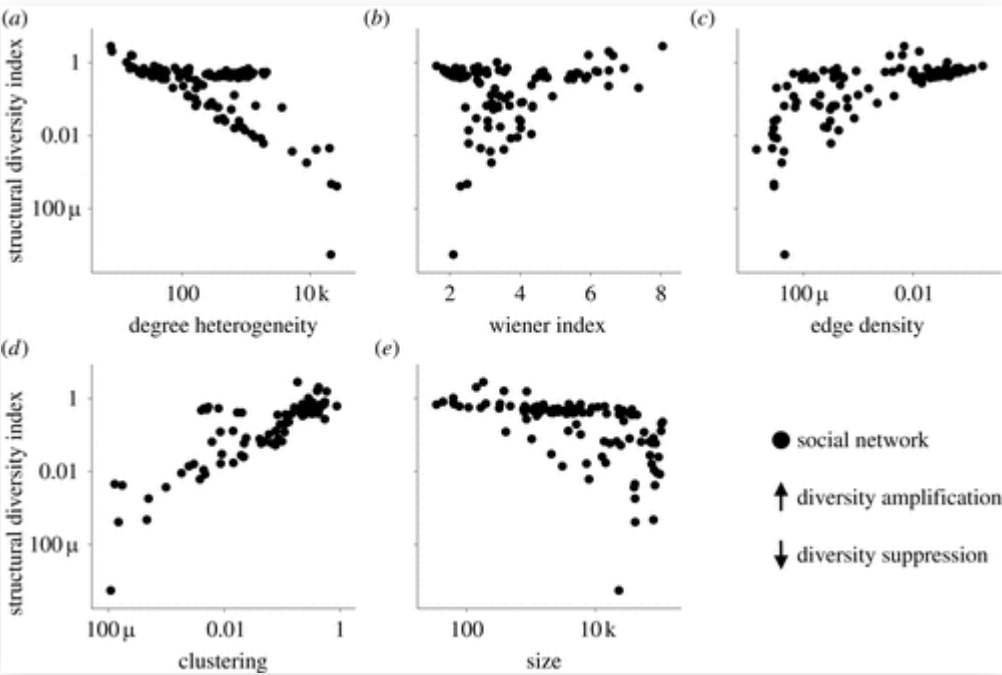


Figure 4. Relationship between the structural diversity index and (a) the degree-heterogeneity $\kappa(G)$, (b) the Wiener index $W(G)$, (c) the edge density $e(G)$, (d) the clustering coefficient $c(G)$ and (e) the size $|V(G)|$ of a network G (see main text for definitions). Analysing a variety of social networks we find that high-degree heterogeneity tends to suppress socio-diversity. By contrast, high clustering, Wiener index and edge density tend to amplify it. The effect of network size is more intricate (see text for details). See Methods for descriptions of the networks.

[Download figure](#) | [Open in new tab](#) | [Download PowerPoint](#)



Table 1.

Dependent variable: Log(structural diversity index). The regression models elucidate the correlations between the structural diversity index and network characteristics as illustrated in figure 4: (i) the correlations with degree-heterogeneity, the Wiener index, and edge density are robust. (ii) The correlation with clustering fades when accounting for degree-heterogeneity, Wiener index, size and edge density. (iii) The correlation with size reverses when factoring in other network characteristics. All variables are standardized for a direct comparison between regression coefficients. See main text for interpretations of these results and Methods for technical details about the regressions.

[View online](#) | [View popup](#)

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.1$.

[Figure 4a](#) reveals a strong negative correlation between the structural diversity index $\Delta(G)$ and degree-heterogeneity $\kappa(G)$, measured as the ratio of the second and first moments of G 's degree distribution [52]. As degree heterogeneity increases, the structural diversity index decreases. This correlation is quite robust. It persists even after accounting for other network properties (see [table 1](#)) and evaluating alternative measures of inequality in the distribution of connection, such as the Gini Index (see Section B of the electronic supplementary material). At its core, this suggests that high-degree heterogeneity tends to suppress socio-diversity. This phenomenon has an intuitive explanation: large-degree vertices ('VIPs', 'hubs', 'influencers' or 'hyperinfluentials' [53]) are crucial—either as initiators or early adopters—in triggering large imitation cascades [53]. This eventually ends up reducing socio-diversity.

Figure 4^b portrays a positive correlation between the structural diversity index $\Delta(G)$ and the Wiener index $W(G)$. Specifically, when the Wiener index has high values, the structural diversity index also tends to be high. The regressions in **table 1** confirm this observation. Moreover, Model 5 in this same table reveals an increase of this correlation when controlling for the effects of other network attributes. From a wider viewpoint, these findings indicate that large network distances between individuals tend to amplify socio-diversity. The reason is intuitive: large distances obstruct meme spreading, a phenomenon testified by the geographical clustering of most cultural forms.

Figure 4^c and Model 3 in **table 1** reveal a positive relationship between the structural diversity index and edge density $e(G)$, defined as the proportion of existing edges to potential edges within the network.

Table 1 (Model 5) demonstrates that this correlation intensifies when accounting for other network characteristics, suggesting that edge density may play an important role in the regulation of socio-diversity. Broadly, in a similar vein to how it fosters meritocracy [54], edge density appears to encourage socio-diversity. These findings align with theories arguing that a greater number of connections reduces the influence of each individual connection, thus diminishing the likelihood of large imitation cascades [12].

Figure 4^d shows a positive correlation between the structural diversity index $\Delta(G)$ and clustering, as measured by the clustering coefficient $c(G)$ [23]. On the surface, higher levels of clustering imply a larger structural diversity index. However, Model 5 of **table 1** highlights that this correlation fades when factoring in other network characteristics. This suggests that these network characteristics might mediate the amplifying effect of clustering on socio-diversity. To break it down, a large network with high average inter-node distances and significant edge density can support socio-diversity irrespective of its clustering levels. However, in our dataset, most dense networks exhibit high clustering. Consequently, it is complicated to disregard the significance of clustering entirely. This viewpoint is further supported by a straightforward mechanism connecting clustering with socio-diversity: clusters, when they are meme-homogeneous, obstruct consensus formation by increasing the persistence of individual memes and decreasing the exposure to new memes.

Finally, **figure 4**^e highlights a negative correlation between the structural diversity index and network size, suggesting that large networks might suppress socio-diversity. However, a closer look (see **table 1**, Model 5), reveals a shift to a positive correlation when factoring in all the discussed network characteristics. This sheds some light on the nuanced relationship between the structural diversity index and network size: Size intrinsically boosts the index, perhaps due to factors like increased overall innovation in larger networks. Yet, as networks grow, they become more sparse because maintaining connections is costly. And this decrease in edge density is likely responsible for the initial negative correlation observed in **figure 4**^e.

3. Discussion

Understanding the interplay between network structure and socio-diversity is crucial, as the latter has numerous positive and negative implications for society. In this article, we have made some steps to understand this. We have found that: (i) A simple index, the structural diversity index, captures the complex interplay between network structure and socio-diversity. (ii) Network characteristics can amplify or suppress socio-diversity: high-degree heterogeneity, as in scale-free networks, tends to suppress it, while high local clustering, large inter-node distances, and significant edge density tend to amplify it. For clarity, we explored the voter model, one of the simplest models of cultural evolution. However, in Section C of the electronic supplementary material, we show that qualitatively similar results hold for other fundamental models such as Axelrod's model [25], Sznaid's model [55] and the (discrete) bounded confidence model [56,57] (see also the review in [58]).

Our work suggests numerous future research directions (see also Section E of the electronic supplementary material). First, an understanding of the consequences that specific characteristics of networks have for socio-diversity implies opportunities for change. For example, our results hint towards possible ways of transforming social networks to sustain greater socio-diversity. For instance, when an increase in degree-heterogeneity causes a reduction in socio-diversity, a simple, decentralized strategy such as ‘stop following the h most connected VIPs in your social network channel’ (or, for short, ‘don’t follow leaders’ [59]) can be surprisingly effective in sustaining it, as shown in [figure 5](#). Future research may explore further kinds of network modification strategies that amplify or suppress socio-diversity. For example, how can one leverage the fact that clustering amplifies socio-diversity?

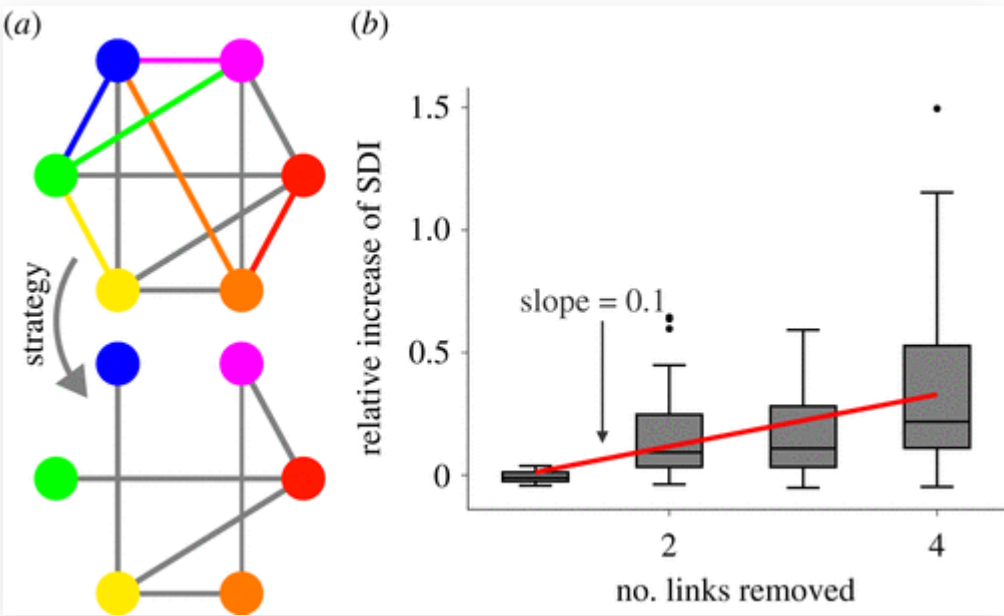


Figure 5. Removing links to highly connected individuals may increase the structural diversity index. (a) Illustration of a simple strategy to raise the structural diversity index: each individual in the network G (top) removes links to her $h = 1$ most connected neighbours (tie sorting is random), resulting in the network G_h (bottom). (b) Percentage change $(\Delta(G_h) - \Delta(G))/\Delta(G)$ in the structural diversity index after applying the procedure outlined in (a) to a network G . This simple procedure leads to remarkable increases of the structural diversity index, even when the number of removed connections per individual is small. On average, increasing h by one leads to a 10% increase in the index. See Methods for details about network data, box plots and regression values.

[Download figure](#) | [Open in new tab](#) | [Download PowerPoint](#)

Second, and most importantly, our findings are primarily rooted in models of cultural evolution, rather than in actual experimental data. It is, therefore, essential to remember that all models inherently simplify the complexities of human interactions. As such, empirical data may reveal nuances that go beyond the narratives presented in our study. Thus, we encourage follow-up research to validate our conclusions through lab or online experiments. In essence, a pivotal question lingers: Can the insights about the relationship between network structure and socio-diversity gained from our modelling and simulations be replicated in an experimental environment?

A third key direction of future research is to deepen our understanding of what levels of socio-diversity are beneficial for distinct social systems. Our work illustrates how one can change social networks to promote or reduce socio-diversity, but it does not tackle whether more or less diversity is desirable. Socio-diversity can bring both advantages and challenges: while an overabundance might lead to division and conflict, too little could hinder innovation and collective intelligence. Pinpointing the ideal balance is complex, necessitating a careful consideration of socio-diversity’s multifaceted effects. Nevertheless, given its profound implications for societal dynamics, it is high time that we appreciate the importance of socio-diversity and explore new ways of shaping it.

4.1. Proof of equation (2.3)

We will present a direct proof of this equation (2.3), based on the fundamentals of random walk theory.

The proof interlinks the model of cultural evolution discussed above with the concept of an r -random walk on a graph G . This r -random walk closely resembles a conventional random walk, but with an added twist: at every step, there is a probability r of the walk ‘halting’. For a more vivid picture, imagine, as Karl Pearson and Lord Rayleigh [19,60], a drunkard wandering through an urban street network. At every intersection, he randomly selects a street and heads towards the next crossing. The nuance in the r -random walk lies here: on any given street, the drunkard might come across a bar he fancies with probability r , leading him to leave the street network indefinitely.

The core relationship between r -random walks and the cultural evolution model discussed above is summarized in the following equation. This equation is based on the principle of ‘voter model duality’ [49]. It is a straightforward generalization of Aldous’ findings for traditional random walks [61]:

$$1 - D(t) = p_r(t). \quad 4.1$$

In this context, $1 - D(t)$ represents the likelihood that, at time step t in the cultural evolution model, a pair of randomly chosen individuals both exhibit the same meme. On the other hand, $p_r(t)$ indicates the probability that two r -random walks, started with uniform probability across the vertices of G , meet before completing t steps.

Equation (4.1) leaves us with the task of understanding $p_r(t)$. Since it simplifies the argument and since our main interest concerns the large-time behaviour of the system, we focus on $p_r(\infty) = \lim_{t \rightarrow \infty} p_r(t)$. This effectively means that we are exploring the probability that two r -random walks on the graph G meet before either one halts.

In mathematical symbols, ‘the probability that two r -random walks on the graph G meet before either one halts’ can be expressed as:

$$p_r(\infty) = P(M_G < \min(S_1, S_2)). \quad 4.2$$

In this equation, M_G is the meeting time of the graph G . This quantity has been defined as the number of steps before two uniformly started *traditional* random walks on G visit the same vertex simultaneously. Meanwhile, S_1 and S_2 are geometric random variables with a success probability r . These random variables count the number of steps taken by the first and second random walks, respectively, before they come to a halt.

To derive equation (2.3), we approximate

$$P(M_G < \min(S_1, S_2)) \approx P(\langle M_G \rangle < \min(S_1, S_2)).$$

Section A of the electronic supplementary material discusses this approximation in detail. Next, since S_1 and S_2 are geometrically distributed random variables with success probability r , $\min(S_1, S_2)$ is a geometrically distributed random variable with success probability $q(r) = 2r - r^2$. Therefore,

$$P(\langle M_G \rangle < \min(S_1, S_2)) = (1 - q(r))^{\langle M_G \rangle} \approx e^{-q(r)\langle M_G \rangle}.$$

^

Back to top

When $r \ll 1$, $q(r) \approx 2r$. Hence, $e^{-q(r)\langle M_G \rangle} \approx e^{-2r\langle M_G \rangle}$. The hypothesis $r \ll 1$ is convenient for presentation because it makes interpretation more straightforward. However, it is not necessary for the paper's conclusions to hold. In fact, replacing $2r$ by $q(r) = 2r - r^2$ yields very similar results.

Drawing upon our prior arguments, we have established

$$p_r(\infty) = P(M_G < \min(S_1, S_2)) \approx e^{-2r\langle M_G \rangle}. \quad 4.3$$

Now, equation (4.1) shows that $1 - D(t) = p_r(t)$. Therefore,

$$p_r(\infty) = \lim_{t \rightarrow \infty} p_r(t) = \lim_{t \rightarrow \infty} 1 - D(t) = 1 - \lim_{t \rightarrow \infty} \sum_{s \leq t} D(s) = 1 - D_\infty. \quad 4.4$$

From this relationship and equation (4.3), it is clear that

$$D_\infty = 1 - p_r(\infty) = 1 - e^{-2r\langle M_G \rangle}. \quad 4.5$$

Finally, replacing the innovation rate r with the *per capita* innovation rate $\alpha = r/|V(G)|$ and recalling that the structural diversity index is defined as the ratio $\Delta(G) = \langle M_G \rangle / |V(G)|$, we obtain the desired equation

$$D_\infty = 1 - e^{-2\alpha\Delta(G)}. \quad 4.6$$

4.2. Structural diversity index of complete, scale-free and Watts–Strogatz networks

For a network G , we defined the structural diversity index by $\Delta(G) = \langle M_G \rangle / |V(G)|$, where M_G is the meeting time of two (uniformly started) random walks on G and $|V(G)|$ is G 's vertex count. Let us discuss estimates for the structural diversity index of complete, scale-free, and Watts–Strogatz networks.

4.2.1. Complete networks

For the complete network K with vertices, the structural diversity index can be computed exactly. In each step, two random walks on K have a probability $1/|V(K)|$ of moving to the same vertex and, hence, of meeting. Consequently, the meeting time M_K of the walks is geometrically distributed with success probability $1/|V(K)|$. In particular, $\langle M_K \rangle = |V(K)|$.

4.2.2. General bounds

For a general network G , exact analytical expressions for the meeting time M_G are not in reach. However, good upper and lower bounds are available.

On the one hand, Cooper *et al.* [62] provide an upper bound for the average meeting time $\langle M_G \rangle$

$$\langle M_G \rangle = O \left(\frac{1}{1 - \lambda_2} \left(2 \log(|V(G)|) + |V(G)| \frac{\langle k \rangle^2}{\langle k^2 \rangle} \right) \right). \quad 4.7$$

Herein, O is the standard asymptotic notation for ‘is asymptotically dominated by’; $\langle k \rangle$ and $\langle k^2 \rangle$ denote, respectively, the first and second moments of G 's degree distribution; and λ_2 indicates the second largest eigenvalue of the transition matrix P of the random walk on G (i.e. the matrix $P(i, j) = 1/\deg(i)$ for all vertices i, j of G).

On the other hand, Aldous [47] provides a lower bound for the average meeting time $\langle M_G \rangle$

$$\langle M_G \rangle = \Omega \left(\frac{|E(G)|}{D_{\max}} \right) \quad 4.8$$

Herein, Ω is the standard asymptotic notation for ‘asymptotically dominates’; D_{\max} is the maximum degree in the graph G ; and $|E(G)|$ is the number of edges.

4.2.3. Scale-free networks

For a scale-free network G_γ with exponent γ , different bounds for $1/(1 - \lambda_2)$ exist, depending on the model with which the network is constructed [63,64]. However, such bounds are typically polynomials in $\log(|V(G_\gamma)|)$. Therefore, neglecting logarithmic terms, equation (4.7) suggests that $\langle M_{G_\gamma} \rangle = O(|V(G_\gamma)| \langle k \rangle^2 / \langle k^2 \rangle)$. Hence, for large networks G_γ , we have

$$\Delta(G_\gamma) \leq \frac{\langle k^2 \rangle}{\langle k \rangle^2}. \quad 4.9$$

Herein, $\langle k \rangle$ and $\langle k^2 \rangle$ denote the first and second moments of G_γ ’s degree distribution. Note that $\langle k \rangle$ and $\langle k^2 \rangle$ depend on the exponent γ . We have calculated their values in terms of $|V(G_\gamma)|$ and γ following the guidelines provided in ch. 6.4 of [52]. This yields the bound for $\Delta(G_\gamma)$ reported in equation (2.4)

$$\Delta(G_\gamma) \leq \begin{cases} 1 & \text{if } \gamma \geq 3 \\ |V(G_\gamma)|^{-((3-\gamma)/(\gamma-1))} & \text{if } 2 < \gamma < 3 \\ |V(G_\gamma)|^{-1} & \text{if } \gamma \leq 2 \end{cases}. \quad 4.10$$

4.2.4. Watts–Strogatz networks

A Watts–Strogatz network W_s with rewiring probability s and average degree $\langle k \rangle$ has $|E(W_s)| = \langle k \rangle |V(W_s)|$ edges and maximum degree $D_{\max} \approx \langle k \rangle$. Therefore, according to equation (4.8), the average meeting time satisfies $\langle M_{W_s} \rangle = \Omega(|V(W_s)|)$. Consequently, for large networks W_s , we retrieve the lower bound reported in equation (2.5)

The following explicit formula captures the relationship between the structural diversity index and the rewiring probability s fairly well

$$\Delta(W_s) \approx \frac{s + 1/\langle k \rangle^2}{s + 1/|V(W_s)|}. \quad 4.12$$

However, we could not find a theoretical derivation of this approximate relationship.

4.3. Network data

All networks have been handled using graph-tool [24] or networkx [65]. The network data employed in this work are described in Section B of the electronic supplementary material, and are freely available at <https://networks.skewed.de>. A Python package enabling the fast numerical computation of the structural diversity index developed by the first author is available through PyPi <https://pypi.org/project/structural-diversity-index/>. See Section D of the electronic supplementary material for more information. ^

4.4. Parameters and specifications for figures

4.4.1. [Figure 2](#)

For selected real-world networks G from our network dataset (see table B.1 and table B.2 of the electronic supplementary material for details), we computed the expected socio-diversity D_∞ by simulating the cultural evolution model with parameter $r = 1/|V(G)|$ for $10 \cdot |V(G)|$ steps 20 times; (ii) the structural diversity index $\Delta(G)$ by simulating 10^4 realizations of M_G and taking the average. Error bars are smaller than the sizes of symbols.

4.4.2. [Figure 3](#)

Simulations were performed on scale-free networks with $N = 10^3$ vertices and minimum degree $m = 4$ and on Watts–Strogatz networks with $N = 10^3$ vertices and average degree $\langle k \rangle = 6$. We computed the structural diversity index $\Delta(G)$ by simulating 10^4 realizations of M_G and averaging. Error bars are smaller than the size of the symbols. The parameters a and b in (A) were obtained by an ordinary least-squares fit of $\log(\Delta(G_\gamma))$ against $\log(|V(G_\gamma)|^{-(3-\gamma)/(\gamma-1)})$. The fit yields $a = 0.3$ and $b = 6.2$ with $R^2 = 0.91$.

4.4.3. [Figure 4](#)

For each real-world network G in our network dataset (see Table B.1 and Table B.2 of the electronic supplementary material for details), we calculated (i) the structural diversity index, (ii) degree-heterogeneity, (iii) Wiener index, (iv) edge density, (v) clustering and (vi) network size. We provide definitions of these quantities in the text and in Section B of the electronic supplementary material. We computed the structural diversity index and Wiener index in the same way as outlined for [figure 2](#); clustering using algorithms from graph-tool [\[24\]](#); degree-heterogeneity, edge density and size by evaluating simple mathematical expressions.

4.4.4. [Figure 5](#)

For selected real-world networks G in our network dataset (see Table B.1 and Table B.2 of the electronic supplementary material for details) and each $h = 1, 2, 3, 4$, we obtained the network G_h through the edge removal procedure described in [figure 5a](#). Specifically, G_h is the largest connected component of the network obtained by the procedure in [figure 5a](#). For each network G_h , we computed the structural diversity index by simulating 10^4 realizations of M_{G_h} and averaging them. Before creating the plot in [figure 5b](#), we cleaned the data by discarding some ‘pathological cases’. First, we discarded all networks with $|V(G_h)| < |V(G)|/4$. These networks were too affected by edge removal for comparisons to be meaningful. Second, we filtered out outliers, i.e. networks such that the relative variation of the structural diversity index $(\Delta(G_h) - \Delta(G))/\Delta(G)$ deviated more than 1 s.d. from the average relative variation of the sample. This procedure discards just a few (about 4) networks for each value of h . The discarded networks all follow a common pattern: the relative variations in their structural diversity indices are anomalously large because of specific structural features. Such anomalous cases are not interesting for our statistical study. This is why we discarded them. The red line is fitted using the ordinary least-squares method ($a = 0.105$, $R^2 = 0.223$).

4.4.5. [Table 1](#)

The coefficients, standard errors and p -values displayed in [table 1](#) are obtained by running ordinary least-square regressions on log-transformed standardized data. Due to space limitations, we excluded the regression analysis of structural diversity against size, as it was considered the least relevant. The work sample used in the regression is the same as that used in [figure 4](#).

4.4.6. Computing realizations of M_G

A realization of M_G is computed by simulating two random walks on the graph G . The simulation is run for $s_{\max} = 100 \cdot |V(G)|$ steps. If the random walks do not meet within s_{\max} steps, we estimate the value of M_G . For this, we use the fact that M_G is approximately geometrically distributed (see Section A of the electronic supplementary material). Specifically, we estimate the geometric distribution that best approximates M_G . Then, we sample from this geometric distribution conditioned on the fact that the sampled value should exceed s_{\max} .

Data accessibility

A Python package enabling the fast numerical computation of the structural diversity index developed by the first author is available through PyPi <https://pypi.org/project/structural-diversity-index/>. A detailed tutorial on how to use the scripts is available on the *eth-coss* public GitHub repository (see <https://github.com/ethz-coss/Structural-diversity-index>) and the full documentation of the scripts can be found on ReadTheDocs (see <https://rse-distance.readthedocs.io/en/latest/>). The network data employed in this study are freely available at <https://networks.skewed.de>.

Supplementary material is available online [66].

Declaration of AI use

We have not used AI-assisted technologies in creating this article.

Authors' contributions

A.M.: conceptualization, formal analysis, methodology, software, visualization, writing—original draft, writing—review and editing; D.H.: conceptualization, funding acquisition, supervision, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration

We declare we have no competing interests.

Funding

This project received financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 833168).

Acknowledgements

A.M. is grateful to Thomas Asikis for support in running simulations. We thank the two anonymous referees for their constructive feedback.

Footnotes

1[↩] The memes considered in this paper should not be confused with the homonymous 'memes' circulating on social media. Social media 'memes' are funny internet content, while this paper's 'memes' are the building blocks of culture.

2 ↩ This is not surprising: ‘Understanding the factors determining biodiversity is arguably the most fundamental problem in ecology and conservation biology’ [4], in contrast to the less central role of socio-diversity studies in many social sciences.

3 ↩ Innovation need not imply a groundbreaking cultural shift but might just be a nuanced variation of an existing meme.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7184003>.

© 2024 The Authors.

Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

Comments

0 Comments

1 Login ▼

G

Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS ?

Name

♥ Share

Best Newest Oldest

Be the first to comment.

Subscribe

Privacy

Do Not Sell My Data

VIEW FULL TEXT | DOWNLOAD PDF

ROYAL SOCIETY OPEN SCIENCE

[About this journal](#)

[Author benefits](#)

[Submit](#)

[Contact information](#)

[Journal metrics](#)

[Search help](#)

ROYAL SOCIETY PUBLISHING

[Our journals](#)

[Historical context](#)

[Open access](#)

[Open science](#)

[Publishing policies](#)

[Permissions](#)

[Conferences](#)

[Videos](#)

[Manage your account](#)

[Terms & conditions](#)

[Privacy policy](#)

[Cookies](#)

THE ROYAL SOCIETY

[About us](#)

[Contact us](#)

[Fellows](#)

[Events](#)

[Grants, schemes & awards](#)

[Topics & policy](#)

[Collections](#)

[Venue hire](#)



THE ROYAL SOCIETY



[Blog](#)

Copyright © 2025 The Royal Society



[Back to top](#)