

LETTER • OPEN ACCESS

## Estimating transmission noise on networks from stationary local order






To cite this article: C. R. Kitching *et al* 2025 *EPL* **150** 31002

View the [article online](#) for updates and enhancements.

You may also like

- [An Algorithm Based on Simple CNN and Bi LSTM Network for Chinese Word Segmentation](#)  
Xiaohan Guan, Xin Liu and Zhi Liu
- [Unveiling the relationship between complex networks metrics and word senses](#)  
Diego R. Amancio, Osvaldo N. Oliveira and Luciano da F. Costa
- [Large language models for causal hypothesis generation in science](#)  
Kai-Hendrik Cohrs, Emiliano Diaz, Vasileios Sitokonstantinou et al.

# Estimating transmission noise on networks from stationary local order

C. R. KITCHING<sup>1(a)</sup> , H. KAUKANEN<sup>2</sup> , J. ABBOTT<sup>1</sup>, D. GOPAL<sup>3</sup> , R. BERMÚDEZ-OTERO<sup>4</sup>   
and T. GALLA<sup>5</sup> 

<sup>1</sup> *Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester  
Manchester M13 9PL, UK*

<sup>2</sup> *Department of Linguistics, University of Konstanz - Universitätsstraße 10, 78464 Konstanz, Germany*

<sup>3</sup> *Institutionen för lingvistik och filologi, Uppsala Universitet - 751 26 Uppsala, Sweden*

<sup>4</sup> *Department of Linguistics and English Language, School of Arts, Languages, and Cultures,  
The University of Manchester - Manchester M13 9PL, UK*

<sup>5</sup> *Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB),  
Campus Universitat Illes Balears - Palma de Mallorca E-07122, Spain*

received 28 February 2025; accepted in final form 7 April 2025  
published online 6 May 2025

**Abstract** – We study networks of nodes characterised by binary traits that change both endogenously and through nearest-neighbour interaction. Our analytical results show that those traits can be ranked according to the noisiness of their transmission using only measures of order in the stationary state. Crucially, this ranking is independent of network topology. As an example, we explain why, in line with a long-standing hypothesis, the relative stability of the structural traits of languages can be estimated from their geospatial distribution. We conjecture that similar inferences may be possible in a more general class of Markovian systems. Consequently, in many empirical domains where longitudinal information is not easily available the propensities of traits to change could be estimated from spatial data alone.

 open access

Copyright © 2025 The author(s)

Published by the EPLA under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

**Introduction.** – Many complex systems can be represented as networks of nodes, each characterised by a set of traits which can change endogenously and through interaction. Examples are abundant in evolutionary biology, gene regulatory systems, and the dynamics of languages [1–4]. The spreading of opinions or diseases are further instances of noisy transmission processes on networks [5,6].

We may naturally wish to estimate the relative propensities of different traits to change. The most direct way is to resort to longitudinal data: in population genetics, for example, cladistics and genetic sequencing have enabled the reconstruction of phylogenetic trees reaching to the beginnings of evolutionary time [7]. More indirect approaches are, however, possible [8]. Mutation rates can also be estimated using mechanistic models together with summary statistics observed in natural populations [7,9–11]. In fact, these indirect methods offer the

<sup>(a)</sup>E-mail: [christopher.kitching@manchester.ac.uk](mailto:christopher.kitching@manchester.ac.uk)  
(corresponding author)

only avenue in fields where access to longitudinal information is restricted: the historical evolution of languages, for instance, can be inferred only to a very shallow temporal depth.

Against this background, Greenberg surmised fifty years ago that the geospatial distribution of the structural traits of languages carries useful information about those traits' temporal dynamics [12]. Motivated by a recent implementation of this intuition [13], we suggest that Greenberg's hypothesis applies to Markovian processes more widely. One can conceive of many such processes as involving the interplay of transmission noise and a tendency to order. We propose that the relative strength of these two, the noise-to-order ratio, can be estimated from a small set of measurable quantities in the stationary state. We call this the *transmission-noise conjecture*.

In this letter we show that the conjecture does indeed hold in networked systems with binary traits that change both spontaneously within nodes and through nearest-

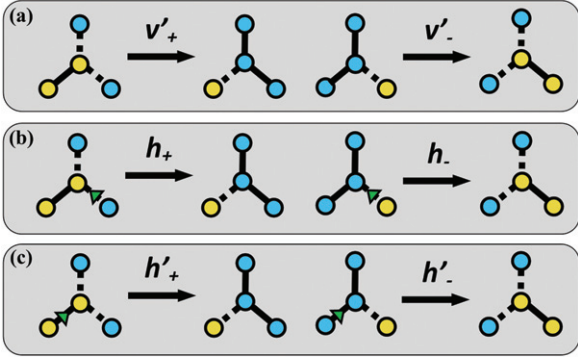


Fig. 1: Illustration of the model dynamics. We show small networks of four nodes which possess (blue) or lack (yellow) a specific trait. Links that connect nodes in the same/different state are solid/dashed. The focal node is the node being considered for update. Panel (a): node changes state endogenously. Panel (b): node correctly copies a neighbour. Panel (c): node erroneously copies a neighbour.

neighbour interaction. Moreover, detailed knowledge of the network is not needed to rank traits according to their noise-to-order ratios.

**Model.** – Our model describes  $N$  nodes  $i$  of an undirected network [14,15]. At time  $t$ , each node is in one of two states, indicated by variables  $s_i(t) = \pm 1$ . In the context of horizontal gene transfer, the binary state could stand for the presence or absence of a mutation in an organism. In a model of opinion dynamics the two states could represent two distinct views on a particular question.

The system evolves through several processes. 1) The state of a node can change endogenously from  $s$  to  $-s$  with rate  $v'_s$ , as shown in fig. 1(a). This reflects transmission errors in time, for example from parent to offspring. 2) A node can change state by faithfully copying the state of a neighbour. The rate coefficients for these events are  $h_{\pm}$  (fig. 1(b)). 3) Finally, a node can change state due to transmission error in an unfaithful copying event (fig. 1(c)); this occurs with rate coefficients  $h'_{\pm}$ . A full definition of the dynamics can be found in the Supplemental Material [Supplementarymaterial.pdf](#) (SM)<sup>1</sup>. The additional references only used in the SM are [16–40].

Our starting point is deliberately general. The events in fig. 1 encompass all networked models with binary states in which nodes can flip either spontaneously or through pairwise interaction with one nearest neighbour. This generalises the model in [13] as well as instances of the so-called “noisy voter model” [41–43]. The subsequent analysis assumes  $h_+ + h'_- = h_- + h'_+$ . This is a mild requirement, indicating that nodes do not preferentially interact with nodes in a particular state (see SM). We stress that, in contrast to most existing literature on the voter model, spontaneous flipping can occur asymmetrically in

<sup>1</sup>The SM contains further details of the model and the theoretical and numerical analysis.

our model ( $v'_+ \neq v'_-$ ), and that the copying of a state from a neighbour can be subject to error ( $h'_+ \neq 0$ ,  $h'_- \neq 0$ ).

**Summary statistics for local order.** – We characterise the stationary state using two summary statistics. One is the trait frequency,  $\rho$ . This is the proportion of nodes which possess that trait (*i.e.*, nodes with  $s_i = +1$ ). The second is the proportion,  $\sigma$ , of links in the network that connect two nodes in opposite states. In line with voter model terminology we will call these active links [44,45].

The average stationary trait frequency is given by

$$\langle \rho \rangle_{\text{st}} = \frac{h'_+ + v'_+}{h'_+ + h'_- + v'_+ + v'_-}. \quad (1)$$

We can show that this expression is valid for any undirected network (see SM).

We also define the ratio

$$\mathbb{H} \equiv \frac{\sigma}{2\rho(1-\rho)}. \quad (2)$$

If, for a given trait frequency  $\rho$ , traits were distributed randomly across the nodes, without correlations, one would have  $\sigma = 2\rho(1-\rho)$ , and thus  $\mathbb{H} = 1$ . When  $\mathbb{H} < 1$  there are fewer active links than random. Thus,  $\mathbb{H}$  quantifies the amount of scatter in the system relative to a random configuration with the same value of  $\rho$ . We note that  $\mathbb{H}$  can take values larger than one for configurations in which the states of neighbouring spins are anti-correlated (see SM).

The typical time course from simulations of the model is illustrated in the  $(\rho, \sigma)$ -plane in fig. 2(a). As indicated by the wiggly lines, the system approaches a parabolic curve of constant  $\mathbb{H}$  quickly, and then fluctuates in the region near the parabola. This has previously been pointed out for voter models in [42,44]. For finite systems, the point defined by the time averages  $\langle \rho \rangle_{\text{st}}$  and  $\langle \sigma \rangle_{\text{st}}$  in the stationary state may not lie on this parabola. This will be discussed further below.

**Transmission noise determines local order.** – Our aim is to demonstrate the validity of the transmission noise conjecture for our model.

As a first step we use the correlation function of spin states at different nodes to show that the stationary density of active interfaces  $\langle \sigma \rangle_{\text{st}}$  can be expressed in terms of the stationary trait frequency  $\langle \rho \rangle_{\text{st}}$  and the parameter combination (see SM)

$$\tau \equiv \frac{(h'_+ + h'_-) + (v'_+ + v'_-)}{\frac{1}{2}[(h_+ + h_-) - (h'_+ + h'_-)]}. \quad (3)$$

This holds for any network. Further, for a wide range of networks we have  $\langle \sigma \rangle_{\text{st}} = 2H(\tau) \langle \rho \rangle_{\text{st}} (1 - \langle \rho \rangle_{\text{st}})$ , where  $H(\tau)$  is dependent upon the network. Importantly, the model parameters enter only through the combination  $\tau$ . In simulations we confirm that  $\langle \mathbb{H} \rangle_{\text{st}} \approx H(\tau)$ , provided  $\tau$  is above some cutoff set by the network size (see SM).

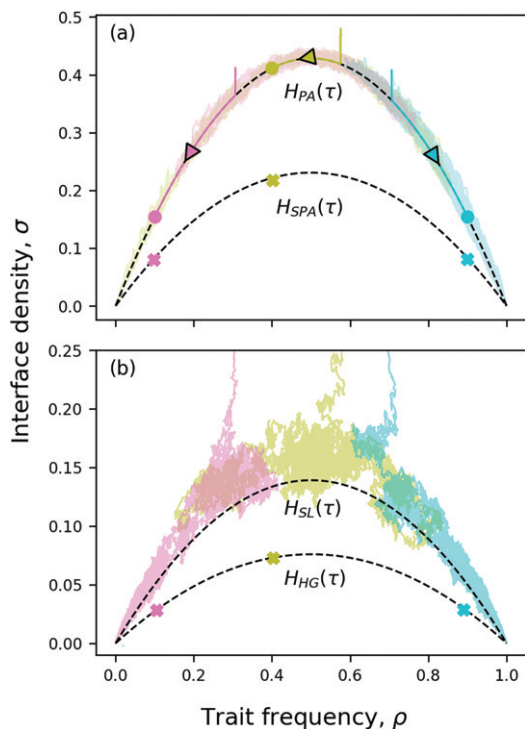


Fig. 2: Model dynamics in the plane spanned by trait frequency  $\rho$  and density of active interfaces  $\sigma$ . (a) is a Barabási-Albert network ( $N = 2500$ ), (b) is the periodic square lattice ( $N = 40000$ ). The wiggly lines show realisations from simulations, each realisation is shown in a different colour. Each wiggly line is for a different set of model parameters, but with a common value of  $\tau = 1 \times 10^{-3}$  in (a), and  $\tau = 1 \times 10^{-4}$  in (b). The solid lines in (a) show the trajectories within the pair approximation, derived in the limit  $N \rightarrow \infty$  (eqs. (S25) and (S42) in the SM). The fixed points of these equations are shown by filled circles. The crosses in each panel indicate the points  $(\langle \rho \rangle_{\text{st}}, \langle \sigma \rangle_{\text{st}})$  for the different parameter sets. Dashed lines are parabolas of constant  $\mathbb{H}$ , obtained from different analytical results, as indicated.

The numerator in eq. (3) describes the average rate of events in the system starting from a fully ordered state ( $\rho = 0$  or  $\rho = 1$ ). The only state changes that are possible in such a configuration are those due to transmission noise. Therefore, the numerator can be interpreted as a strength of transmission noise. The denominator in the expression for  $\tau$  is the rate with which correlations build up between neighbouring spins, starting from a random initial condition with  $\rho = 1/2$  (see SM). Thus, we can think of  $\tau$  as a noise-to-order ratio, and hence a measure of instability<sup>2</sup>.

This interpretation, together with the observation that local order is set by  $\tau$  (*i.e.*,  $H$  is a function of  $\tau$ ), confirms the transmission noise conjecture in our model.

**Inference of  $\tau$  from observed data: analytical results.** – We now turn to the question of inferring the

<sup>2</sup>The quantity  $\tau$  can take negative values. This describes systems with anti-correlations between neighbouring spins. For details see sect. S8 B in the SM.

noise-to-order ratio  $\tau$ . This is possible if  $\langle \mathbb{H} \rangle_{\text{st}}$  can be estimated from observations, and, if the functional form of  $H(\tau)$  can be calculated either exactly or as an approximation. We now discuss how to do the latter, and then turn to an empirical example.

The pair approximation (PA) is a standard tool for the analysis of interacting dynamics on infinite uncorrelated networks [44,46]. Further, the so-called stochastic pair approximation (SPA) captures leading-order corrections in the network size [43]. For our model we find (see SM)

$$H_{\text{PA}}(\tau) = \frac{\mu - 2 - \mu\tau + \sqrt{(\tau + 1)^2\mu^2 - 4(\mu - 1)}}{2(\mu - 1)}, \quad (4)$$

and

$$H_{\text{SPA}}(\tau) = H_{\text{PA}}(\tau) \left[ 1 - \frac{\mu_2}{N\mu^2} \frac{\tau + H_{\text{PA}}(\tau)}{\tau} \right] + \mathcal{O}(N^{-2}), \quad (5)$$

where  $\mu$  is the mean degree of the network and  $\mu_2$  is the second moment of the degree distribution.

In addition to these approximations, we can calculate  $H(\tau)$  exactly for networks which have the following properties for all fixed integers  $\ell$ : The number of distinct walks of length  $\ell$  starting at any node is the same. Additionally, the fraction  $\Omega^{(\ell)}$  of those walks ending at the starting point is also the same for all nodes. For such “homogeneous” networks (HG) we show that (see SM)

$$H_{\text{HG}}(\tau) = \frac{1 + \tau}{\sum_{\ell=0}^{\infty} \Omega^{(\ell)} / (\tau + 1)^\ell} - \tau. \quad (6)$$

Simulations show that this expression works well when the degree distribution is sufficiently tight, even if the network is not strictly homogeneous (see SM).

The coefficients  $\Omega^{(\ell)}$  can either be calculated in closed form (see the SM for examples), or be obtained by direct enumeration of walks.

**Interpretation and test in simulations.** – Figure 2(a) illustrates the relation between the different approximations for Barabási-Albert networks [47]. We show simulations for different parameter combinations, but keeping  $\tau$  in eq. (3) fixed. The trajectories in the  $(\rho, \sigma)$ -space obtained from the PA for infinite systems converge to fixed points on the parabola  $\sigma = 2H_{\text{PA}}(\tau)\rho(1 - \rho)$ . Trajectories from simulations of finite systems fluctuate about these fixed points and remain near the parabola. The time averages in the stationary state,  $\langle \rho \rangle_{\text{st}}$  and  $\langle \sigma \rangle_{\text{st}}$ , are indicated by crosses in fig. 2(a), and lie below the parabola set by  $H_{\text{PA}}(\tau)$ . This deviation is a consequence of finite-size fluctuations. As seen in the figure, the SPA captures this effect to good accuracy.

A similar effect is found for square lattices (fig. 2(b)). Trajectories for finite systems fluctuate near the parabola  $\sigma = 2H_{\text{SL}}(\tau)\rho(1 - \rho)$  (originally derived in [13], where an analytical formula for  $H_{\text{SL}}(\tau)$  was given). The time average  $(\langle \rho \rangle_{\text{st}}, \langle \sigma \rangle_{\text{st}})$ , however, is not on this parabola, but is captured by eq. (6).

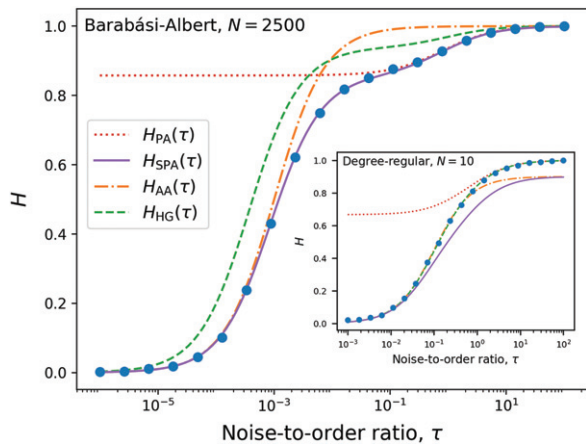


Fig. 3: Order parameter  $H(\tau)$  within the different theories discussed in the text. Main axis: Barabási-Albert network ( $N = 2500$ , mean degree  $\mu \approx 8$ ). Inset: degree-regular network ( $N = 10$ ,  $\mu = 4$ ). Lines show the analytical results as indicated. Blue markers are from simulations, averaged over 100–10000 independent samples in the steady state. Model parameters for a given choice of  $\tau$  can be randomly generated via the algorithm described in the SM.  $H_{AA}(\tau)$  is the “annealed approximation” (see SM).

We test the different approximations further in fig. 3, focusing on Barabási-Albert networks of size  $N = 2500$ , and a small random degree-regular network ( $N = 10$ ). The PA fails for  $\tau \rightarrow 0$ , because complete ordering can only occur on finite networks. We find  $\lim_{\tau \rightarrow 0} H_{PA} = \frac{\mu-2}{\mu-1}$ , describing long-lived partially ordered states of the voter model [44]. For the Barabási-Albert network the assumption of homogeneity is not valid and eq. (6) is inaccurate. The SPA, on the contrary, describes simulations well in the main panel, but becomes inaccurate for large  $\tau$  on small networks (inset) due to finite-size effects beyond leading order. For random degree-regular networks (inset) the HG result in eq. (6), on the other hand, produces satisfactory results for all  $\tau$ . We have deliberately chosen a small network to demonstrate that the HG approach does not require the network to be large.

**Stability of language traits.** – The World Atlas of Language Structures (WALS) [48] reports geographical coordinates and information on language traits for 2662 of the approximately 7000 existing human languages. We extract 35 binary (or binarisable) traits, such as the presence or absence of a definiteness marker (*e.g.*, the English definite article *the*).

For each trait, we measure the fraction of languages possessing the trait,  $\rho$ . We use geographic distance between languages to construct a putative interaction network. Each node is a language and the processes operated on the node represent a coarse-grained view of the dynamics of speaker behaviour (*e.g.*, acquisition of the language by children, or the use of traits of a neighbouring language by bilingual speakers). Further details can be found in the

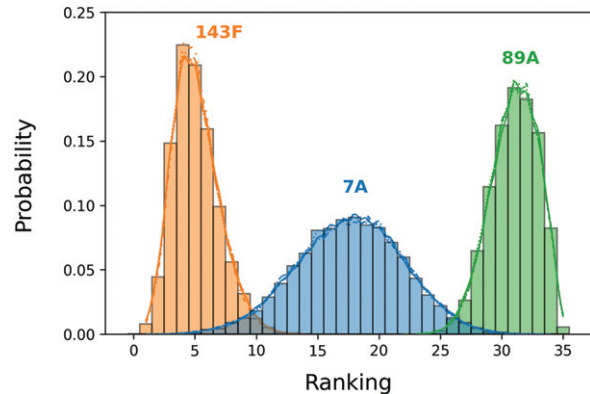


Fig. 4: Rankograms for three linguistic traits 143F, 7A, and 89A from [48], each obtained from 10000 bootstrap samples. For each trait we show rankograms obtained from  $\mathbb{H}$  directly, and from  $\tau$  obtained using the functions  $H_{AA}(\tau)$ ,  $H_{SL}(\tau)$ ,  $H_{HG}(\tau)$  and  $H_{SPA}(\tau)$ . These are virtually indistinguishable from each other on the network. For clarity we show Gaussian kernel density estimates, for  $H_{HG}$  we show the full histogram. Further details are given in the SM along with rankograms for the remaining traits.

SM. Using this network we calculate the interface density  $\sigma$ . To assess the uncertainty due to the incompleteness of WALS we use bootstrap sampling (*i.e.*, we sample subsets of languages from WALS). This allows us to obtain a distribution for  $\mathbb{H} = \sigma/[2\rho(1-\rho)]$  for each trait. Further details of the procedure can be found in the SM.

In ref. [13] the values for  $\mathbb{H}$  were converted into  $\tau$  by inverting the formula  $H_{SL}(\tau)$  for the infinite square lattice. This was done without proper justification, as no other analytical solution for  $H(\tau)$  was available. Surprisingly, the estimates of  $\tau$  constructed in this manner correlate well with other measures of the (in)stability of language traits [13,49]. This agreement, while welcome, was somewhat puzzling.

The results of the present work explain these observations. We can use the different functional forms  $H_X(\tau)$  to convert the distributions of  $\mathbb{H}$  into distributions of  $\tau_X$ . The subscript  $X$  denotes the different theories SPA, AA, SL or HG. For a given  $X$ , we can use the distributions of  $\tau_X$  to construct rankograms for the language traits. The rankogram of each trait describes the probability that the trait is ranked  $r$ -th. We illustrate this in fig. 4. The width of a trait’s rankogram reflects the confidence we can have in its relative instability based on the data. In [13] the ranking was based on median values of  $\tau$  only.

Strikingly, the rankogram for any given trait is the same irrespective of the choice  $X$  made to infer  $\tau$ . That is to say, no matter which of the analytical forms  $H_X(\tau)$  we use to obtain  $\tau$  from  $\mathbb{H}$ , the ranking of traits will remain the same. Noting that  $H_{PA}(\tau)$  and  $H_{SPA}(\tau)$  only use the first two moments of the empirical degree distribution and that the different approximations  $H_X(\tau)$  assume different network types, this ultimately means that the detailed



topology of the interaction network is not important. This robustness explains the surprising success of the stability estimates in [13] obtained using the analytical solution of the model on a square lattice. Figure 4 also shows that ranking on  $\mathbb{H}$  itself (*i.e.*, based on the amount of spatial scatter) gives similar results, thus confirming Greenberg’s hypothesis that low scattering indicates relative stability, whereas highly scattered traits would be comparatively unstable.

**Discussion.** – This letter has demonstrated the transmission-noise conjecture for two-state Markovian dynamics with nearest-neighbour interaction on networks. Traits can be ranked by stability from purely spatial information; neither longitudinal data nor detailed knowledge of the interaction network are required. We have used these results to put a long-standing hypothesis about the stability of linguistic traits on firm mathematical ground.

While we have explained the success of the *ad hoc* method of [13], our analysis remains subject to limitations. For example, the model in its current form does not capture temporal networks [50], dynamics with absorbing states [5,51], or cyclic behaviour and frequency-dependent selection in evolutionary games [52–56].

Despite these limitations our model has proved able to recover important facts about the temporal dynamics of language from purely spatial data. We believe that similar applications are possible in the study of opinion change, using, for example, the dataset in [57], or of a range of cases in cultural evolution for which spatial data are more readily available than longitudinal data: the evolution of ethnographic traits in anthropology [58,59], the evolution of musical styles [60–62], and further afield, bird and whale song [63–65]. In addition, further theoretical work should explore how generally it is the case that the noise-to-order ratio in the dynamics of a complex system can be inferred from configurations in the stationary state.

\*\*\*

We acknowledge support from the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (FEDER, UE) under project APASOS (PID2021-122256NB-C21, PID2021-122256NB-C22), the María de Maeztu programme for Units of Excellence, CEX2021-001164-M funded by MCIN/AEI/10.13039/501100011033, ERC grant No. 851423, and the Government of the Balearic Islands CAIB, fund ITS2017-006 under project CAFECONMIEL (PDR2020/51). We further acknowledge an EPSRC studentship (EP/T517823/1), and support by the Royal Society UK (APEX Award APX\R1\211253). We thank ANTONIO FERNÁNDEZ PERALTA for useful discussions regarding the stochastic pair approximation.

*Data availability statement:* All data that support the findings of this study are included within the article (and any supplementary files).

## REFERENCES

- [1] HUSON D. H. and BRYANT D., *Mol. Biol. Evolution*, **23** (2005) 254.
- [2] ALON U., *Nat. Rev. Genet.*, **8** (2007) 450.
- [3] BECKNER C., BLYTHE R., BYBEE J., CHRISTIANSEN M. H., CROFT W., ELLIS N. C., HOLLAND J., KE J., LARSEN-FREEMAN D. and SCHOENEMANN T., *Language Learning*, **59** (2009) 1.
- [4] BAXTER G. J., BLYTHE R. A. and MCKANE A. J., *Phys. Rev. Lett.*, **101** (2008) 258701.
- [5] CASTELLANO C., FORTUNATO S. and LORETO V., *Rev. Mod. Phys.*, **81** (2009) 591.
- [6] KISS I., MILLER J. C. and SIMON P. L., *Mathematics of Epidemics on Networks From Exact to Approximate Models* (Springer International Publishing, New York, NY) 2017.
- [7] HEY J. and MACHADO C. A., *Nat. Rev. Genet.*, **4** (2003) 535.
- [8] DETTMER S. L. and BERG J., *J. Stat. Mech.: Theory Exp.*, **2018** (2018) 023403.
- [9] NACHMAN M. W. and CROWELL S. L., *Genetics*, **156** (2000) 297.
- [10] FOSTER P. L., *Methods for determining spontaneous mutation rates in DNA Repair, Part B, Methods in Enzymology*, Vol. **409** (Academic Press) 2006, pp. 195–213.
- [11] ROSCHE W. A. and FOSTER P. L., *Methods*, **20** (2000) 4.
- [12] GREENBERG J. H., *Language Typology: A Historical and Analytic Overview* (De Gruyter Mouton, Berlin, Germany) 1974.
- [13] KAUFANEN H., GOPAL D., GALLA T. and BERMÚDEZ-OTERO R., *Sci. Adv.*, **7** (2021) eabe6540.
- [14] NEWMAN M. E. J., *Networks* (Oxford University Press, Oxford, UK) 2018.
- [15] BARABÁSI A.-L., *Network Science* (Cambridge University Press, Cambridge, UK) 2016.
- [16] BERMÚDEZ-OTERO R., *Glossa: J. Gen. Linguist.*, **5** (2020) 1.
- [17] HICKEY R., *The Handbook of Language Contact* (Blackwell Publishing Ltd, Hoboken, NJ, USA) 2010.
- [18] COMRIE B., DRYER M. S., GIL D. and HASPELMATH M., *Locations of languages (v2020.4)*, in *The World Atlas of Language Structures Online* (Zenodo) 2013, [https://wals.info/chapter/s1#3.4.Locations\\_of\\_languages](https://wals.info/chapter/s1#3.4.Locations_of_languages).
- [19] JOHNSON R. W., *Teaching Stat.*, **23** (2001) 49.
- [20] BOSLAUGH S., *Statistics in a Nutshell: A Desktop Quick Reference* (O’Reilly Media, Inc., Sebastopol, Cal., USA) 2012.
- [21] DEDIU D. and CYSOUW M., *PLoS ONE*, **8** (2013) e55009.
- [22] COMMANDER C. W., *Maximum cut problem, max-cut*, in *Encyclopedia of Optimization*, edited by FLOUDAS C. A. and PARDALOS P. M. (Springer US, Boston, Mass.) 2009, pp. 1991–1999.
- [23] SALANTI G., ADES A. E. and IOANNIDIS J. P. A., *J. Clin. Epidemiol.*, **64** (2011) 163.
- [24] OLDHAM K., MYLAND J. and SPANIER J., *An Atlas of Functions* (Springer, New York, NY) 2008.
- [25] SATO K. and IWASA Y., *Pair approximations for lattice-based ecological models*, in *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, edited by DIECKMANN U., LAW R. and METZ J. A. J., *Cam-*

- bridge Studies in Adaptive Dynamics* (Cambridge University Press, Cambridge, UK) 2000, pp. 341–358.
- [26] STANLEY R. P., *Catalan Numbers* (Cambridge University Press, Cambridge, UK) 2015.
- [27] WANLESS I. M., *Combinator. Probab. Comput.*, **19** (2010) 463.
- [28] BORWEIN J. M. and BORWEIN P. B., *Pi and the AGM: A Study in the Analytic Number Theory and Computational Complexity* (John Wiley & Sons, Inc, Hoboken, NJ, USA) 1987, [https://wayback.cecm.sfu.ca/~pborwein/TEMP\\_PROTECTED/pi-agm.pdf](https://wayback.cecm.sfu.ca/~pborwein/TEMP_PROTECTED/pi-agm.pdf).
- [29] KNUTH D. E., *The art of computer programming*, 3rd edition, Vol. 1: *Fundamental Algorithms* (Addison Wesley Longman, Boston, Mass.) 1997, [https://haio.ir/wp-content/uploads/2024/12/Donald-Knuth-The-Art-of-Computer-Programming-Vol.-1\[haio.ir\]-Fundamental-Algorithms-3rd-Edition-Addison-Wesley-Professional-1997.pdf](https://haio.ir/wp-content/uploads/2024/12/Donald-Knuth-The-Art-of-Computer-Programming-Vol.-1[haio.ir]-Fundamental-Algorithms-3rd-Edition-Addison-Wesley-Professional-1997.pdf).
- [30] ABBOTT J., *Laplace transforms of products of modified Bessel functions* (March 31 2022), <https://math.stackexchange.com/questions/4404226/laplace-transforms-of-products-of-modified-bessel-functions>.
- [31] MATSUMOTO K., *Appell and lauricella hypergeometric functions*, in *Encyclopedia of Special Functions: The Askey-Bateman Project*, edited by KOORNWINDER T. H. and STOKMAN J. V. (Cambridge University Press, Cambridge, UK) 2020, pp. 79–100.
- [32] LAURICELLA G., *Rendiconti del Circolo Matematico di Palermo*, **7** (1893) 111.
- [33] PRUDNIKOV A. P., BRYCHKOV Y. A. and MARICHEV O. I., *Integrals and Series Volume 4: Direct Laplace Transforms*, Vol. 4 (Gordon and Breach Science Publishers, Reading, UK) 1992.
- [34] SCHMIDT M. D., *Axioms*, **8** (2019) 62.
- [35] FLAJOLET P. and SEDGEWICK R., *Analytic Combinatorics* (Cambridge University Press) 2009.
- [36] GRAHAM R. L., KNUTH D. E., PATASHNIK O. and LIU S., *Comput. Phys.*, **3** (1989) 106.
- [37] SLOANE N. J. A., *The on-line encyclopedia of integer sequences*, <http://oeis.org>.
- [38] GLEESON J. P., *Phys. Rev. X*, **3** (2013) 021004.
- [39] DOROGOVTSSEV S., *Lectures on Complex Networks* (Oxford University Press, Oxford, UK) 2010.
- [40] KNOBLAUCH A., *SIAM J. Appl. Math.*, **69** (2008) 197.
- [41] GRANOVSKY B. L. and MADRAS N., *Stoch. Processes Appl.*, **55** (1995) 23.
- [42] CARRO A., TORAL R. and MIGUEL M. S., *Sci. Rep.*, **6** (2016) 1.
- [43] PERALTA A. F., CARRO A., MIGUEL M. S. and TORAL R., *New J. Phys.*, **20** (2018) 103045.
- [44] VAZQUEZ F. and EGUÍLUZ V. M., *New J. Phys.*, **10** (2008) 063011.
- [45] REDNER S., *C. R. Phys.*, **20** (2019) 275.
- [46] PUGLIESE E. and CASTELLANO C., *EPL*, **88** (2009) 58004.
- [47] BARABÁSI A.-L. and ALBERT R., *Science*, **286** (1999) 509.
- [48] DRYER M. S. and HASPELMATH M., (Editors) *WALS Online (v2020.4)* (Zenodo) 2013, <https://wals.info/>.
- [49] DEDIU D., *Proc. R. Soc. B: Bio. Sci.*, **278** (2011) 474.
- [50] HOLME P. and SARAMÄKI J., *Phys. Rep.*, **519** (2012) 97.
- [51] AL HAMMAL O., CHATÉ H., DORNIC I. and MUÑOZ M. A., *Phys. Rev. Lett.*, **94** (2005) 230601.
- [52] TRAUlsen A. and HAUERT C., *Stochastic Evolutionary Game Dynamics* (John Wiley & Sons, Ltd, Hoboken, NJ, USA) 2009, Chapt. 2, pp. 25–61.
- [53] AHERN C. and CLARK R., *Semantics and Pragmatics*, **10** (2017) 11.
- [54] KAUMANEN H., *R. Soc. Open Sci.*, **7** (2020) 201682.
- [55] BAUMANN A. and RITT N., *Phonology*, **34** (2017) 439.
- [56] DEO A., *Seman. Pragmat.*, **8** (2015) 1.
- [57] HOFMANN V., SCHÜTZE H. and PIERREHUMBERT J. B., *The Reddit Politosphere: a large-scale text and network resource of online political discourse*, in *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. **16** (AAAI) 2022, pp. 1259–1267.
- [58] MURDOCK G. P., *Atlas of World Cultures* (University of Pittsburg Press, Pittsburg, Penn., USA) 1981.
- [59] GOODY J., *World Ethnographic Atlas, 1967–1971* (1974), UK Data Service.
- [60] LOMAX A. and BERKOWITZ N., *Science*, **177** (1972) 228.
- [61] SAVAGE P. E., *Palgrave Commun.*, **5** (2019) 1.
- [62] VERHOEF T. and RAVIGNANI A., *Front. Psychol.*, **12** (2021) 668300.
- [63] SLATER P. J. B., *Trends Ecol. Evol.*, **1** (1986) 94.
- [64] APLIN L. M., *Animal Behaviour*, **147** (2019) 179.
- [65] GARLAND E. C., GOLDIZEN A. W., REKDAHL M. L., CONSTANTINE R., GARRIGUE C., HAUSER N. D., POOLE M. M., ROBBINS J. and NOAD M. J., *Curr. Biol.*, **21** (2011) 687.

# Estimating transmission noise on networks from stationary local order

## —SUPPLEMENTAL MATERIAL—

Christopher R. Kitching,<sup>1,\*</sup> Henri Kauhanen,<sup>2,†</sup> Jordan Abbott,<sup>1,‡</sup>  
Deepthi Gopal,<sup>3,§</sup> Ricardo Bermúdez-Otero,<sup>4,¶</sup> and Tobias Galla<sup>5,\*\*</sup>

<sup>1</sup>*Department of Physics and Astronomy, School of Natural Sciences,  
The University of Manchester, Manchester M13 9PL, UK*

<sup>2</sup>*Department of Linguistics, University of Konstanz,  
Universitätsstraße 10, 78464 Konstanz, Germany*

<sup>3</sup>*Institutionen för lingvistik och filologi, Uppsala Universitet, 751 26 Uppsala, Sweden*

<sup>4</sup>*Department of Linguistics and English Language; School of Arts, Languages,  
and Cultures; The University of Manchester, Manchester M13 9PL, UK*

<sup>5</sup>*Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB),  
Campus Universitat Illes Balears, E-07122 Palma de Mallorca, Spain*

(Dated: February 18, 2025)

### Contents

S1. Full model definition and notation	S3
A. Model definition	S3
1. Microscopic configurations of the model	S3
2. Trait frequency and density of active links	S3
3. Model dynamics	S4
B. Notation	S5
C. Relation of the present setup to the model in [S1]	S5
S2. Complete network	S7
A. Transition rates in continuous-time and master equation	S7
B. Magnetisation	S7
C. Interface density	S9
S3. Pair approximation	S9
A. Transition rates for individual nodes	S9
B. Magnetisation	S10
C. Interface density	S10
D. Further discussion of the function $H_{\text{PA}}(\tau)$ obtained from the pair approximation	S12
1. Regions of validity of the two branches of the solution	S12
2. Comparison against simulations	S12
3. Limits on the inference of $\tau$ from $H(\tau)$	S13
E. Limit of all-to-all interaction	S13
S4. Annealed approximation	S14
A. Nature of the approximation and the rate equations	S14
B. Master equation	S15
C. Magnetisation	S15
D. Interface density	S18
E. Comparison against simulations	S21
S5. Analytical treatment of the model on an infinite square lattice	S22

---

\* christopher.kitching@manchester.ac.uk

† henri.kauhanen@uni-konstanz.de

‡ jordanabbott2013@hotmail.co.uk

§ deepthi.gopal@lingfil.uu.se

¶ ricardo.bermudez-otero@manchester.ac.uk

\*\* tobias.galla@ifisc.uib-csic.es



	S2
A. Setup and spin flip rates	S22
B. Magnetisation	S23
C. Interface density	S23
S6. Approach based on network walks	S25
A. Useful definitions and identities for the further analysis	S25
1. Definitions	S25
2. Identities	S26
B. Magnetisation	S28
C. Generalisation to ‘weighted magnetisation’	S30
D. Interface density	S31
E. Homogeneous networks	S34
1. Homogeneity assumptions	S34
2. Interface density for homogeneous networks, and resulting simplifications	S35
3. First special case: Finite complete networks	S36
4. Second special case: Infinite hyper-cubic lattice in $d$ -dimensions	S37
5. Application to infinite two-dimensional square lattice	S39
6. Third special case: Infinite Bethe lattice	S40
7. Calculation of $\Omega^{(\ell)}$ and validity of the method	S42
S7. Stochastic pair approximation	S45
A. Master equation for reduced set of degrees of freedom	S45
B. Dynamics of first and second moments	S46
C. System-size expansion	S48
S8. Scattering and the noise-to-order ratio	S52
A. Relation between $\mathbb{H}$ and $H(\tau)$	S52
1. Degree of scatter for individual realisations	S52
2. Relation between average density of active interfaces and average magnetisation	S52
3. $H(\tau)$ can be inferred from measurement $\mathbb{H}$ in individual configurations, for large $\tau$	S53
4. Range of possible $H(\tau)$ values	S54
B. Noise-to-order ratio, $\tau$	S54
1. Definition and interpretation	S54
2. Range of possible values for $\tau$	S55
C. Behaviour in different regimes	S56
1. The limit $\tau \rightarrow -\infty$	S56
2. Finite negative values of $\tau < -2$	S56
3. The limit $\tau \rightarrow -2^-$	S57
4. The limit $\tau \rightarrow 0^+$	S57
5. Finite $\tau > 0$	S57
6. The limit $\tau \rightarrow \infty$	S57
7. Illustration and test against simulations	S58
S9. Analysis of data from the World Atlas of Language Structures	S59
A. Further details on the speaker-level processes underpinning the model	S59
B. Dataset	S60
C. Empirical noise-to-order ratio	S60
D. Comparison against longitudinal stability estimate	S61
E. Bootstrapping and rankograms	S63
1. Construction of rankograms	S63
2. Distributions of $\tau_X$	S65
F. SUCRA scores and comparison with Dediu	S66
S10. Further mathematical background	S67
A. Exponential generating function for modified Bessel functions of the first kind	S67
B. Determining standard model parameters for a given $\tau$ and $\langle m \rangle_{\text{st}}$	S68
1. The case $\tau > 0$	S68
2. The case $\tau = 0$ (standard voter model)	S69
3. Negative values of $\tau$ ( $\tau \leq -2$ )	S69
References	S70

## Overview

In this supplement we provide the technical details of the results presented in the main paper. This document is structured as follows:

Firstly, in Sec. S1 we give a full definition of the model and some identities that will be used throughout the supplement in order to simplify the calculations.

The next several sections focus on deriving equations for the steady-state magnetisation and interface density on various topologies. In Sec. S2 we focus on the model on complete networks. In Sec. S3 we then study the model on infinite uncorrelated networks using the pair approximation. This is where the noise-to-order ratio,  $\tau$ , is first introduced. In Sec. S4 we then look at the model on finite uncorrelated networks using the annealed approximation. Sec. S5 focuses on the model on an infinite square lattice.

In Sec. S6 we present an approach based on counting walks. We are able to derive the steady-state magnetisation exactly for any network, and the steady-state interface density for homogeneous networks of any size. We present closed form solutions for a number of specific networks such as finite size complete networks,  $d$ -dimensional lattices, and infinite regular trees.

In Sec. S7 we use the stochastic pair approximation for our model, improving on the conventional pair approximation in Sec. S3. This is valid for finite uncorrelated networks.

In Sec. S8 we discuss the interpretations of the quantities  $\mathbb{H}$ ,  $H(\tau)$  and  $\tau$ , and the connection between them. This includes an analysis of cases with negative  $\tau$ .

In Sec. S9 we go into detail on how the theory can be applied to linguistic data. This includes details about the bootstrapping method and ranking statistics that are presented in the main paper.

Finally, Sec. S10 contains miscellaneous proofs and further details of numerical algorithms.

### S1. Full model definition and notation

#### A. Model definition

##### 1. Microscopic configurations of the model

Throughout, we use the mathematical notion of networks, i.e. nodes connected by links which determine which nodes can interact. We only consider undirected networks. In cases where we are considering a single network we will often use  $P_k$  to denote the frequency of nodes which have degree  $k$ , and then refer to this as ‘the degree distribution’. However, we note that individual members of a network ensemble may have empirical degree frequencies different from the degree distribution in the ensemble as a whole. Further, the term ‘degree-regular’ network means networks where all nodes have the same degree. In the context of simulations we specifically used random degree-regular networks. Networks where all nodes have the same degree, but have some further structure, will be referred to by their specific names, e.g. the 2D infinite square lattice or the Bethe lattice.

We consider various ‘traits’, and the nodes in the network can either process or lack a given trait. In this way, we assign node  $i$  a property  $s_i \in \{-1, 1\}$ . We will use physics terminology and call this a *spin*. A node with  $s_i = +1$  is referred to as ‘spin-up’ or the node being in the ‘up-state’. This means that the node has the particular trait that we are considering. A node with  $s_i = -1$ , referred to as ‘spin-down’ or the node being in the ‘down-state’, means that the node lacks the particular trait we are considering. It is important to note that we only consider a single trait at any one time, there is no interaction between traits.

##### 2. Trait frequency and density of active links

We adopt the voter model terminology in that links connecting nodes in states  $s = +1$  and  $s = -1$  nodes are referred to as *active* links. Links that connect same spin nodes, i.e.  $s = \pm 1$  to  $s = \pm 1$  nodes are referred to as

*inactive* links. This convention from the voter model comes from the fact that the only existing processes in the voter model are error-free copying processes between neighbours. These can only occur between opposite spin nodes, hence those links are called active [S2].

The two quantities of interest are the *trait frequency*,  $\rho$ , and *interface density*,  $\sigma$ . The trait frequency is defined as the proportion of nodes in state  $s = +1$ . Sometimes it is mathematically simpler to work in terms of the *magnetisation*,  $m$ , defined as  $m = \frac{1}{N} \sum_i s_i$ , where we are summing over all nodes in the network. The magnetisation and trait frequency are related via  $m = 2\rho - 1$ . The interface density  $\sigma$  is defined as the proportion of active links among all links in the network.

### 3. Model dynamics

There are two ‘modes’ in which the dynamics can act:

1. endogenous changes of a node,
2. changes due to interactions between nodes.

The first is modelled by the spontaneous flipping of a node’s state. This is what is referred to as ‘noise’ in conventional noisy voter models [S3, S4]. However, in our model the rates for spontaneous flips from  $+1$  to  $-1$  and those in the reverse direction may differ from each other. In [S1] endogenous processes are referred to as ‘vertical’ events. In our model any vertical event leads to the state change of one spin, that is the spontaneous acquisition or loss of a trait in that node. This can be understood as ‘unfaithful’ vertical transmission from one generation to the next, i.e., the next generation of the entity represented by the node incorrectly copies the state of the previous generation at that node. We do not include faithful vertical events, as these do not lead to any state changes.

In the second mode a spin may change state due to an interaction with a neighbouring node. This will be detailed below. In [S1] such processes are referred to as ‘horizontal’ events. A copying process from a neighbour can be *faithful* or *unfaithful*. Faithful horizontal transmission means that the node correctly copies the state of the neighbour. Unfaithful horizontal transmission means that an error occurs in a copying event, so that one node adopts the state opposite to that of the neighbour it attempts to copy.

Six model parameters are needed to define the above dynamics. They are defined as follows (rates associated with unfaithful events are indicated by a dash):

$$\begin{aligned}
 v'_+ &= \text{rate with which a node gains the trait through} \\
 &\quad \text{unfaithful vertical transmission,} \\
 v'_- &= \text{rate with which a node loses the trait through} \\
 &\quad \text{unfaithful vertical transmission,} \\
 h_+ &= \text{rate coefficient for events in which a node} \\
 &\quad \text{gains the trait through faithful horizontal transmission,} \\
 h_- &= \text{rate coefficient for events in which a node} \\
 &\quad \text{loses the trait through faithful horizontal transmission,} \\
 h'_+ &= \text{rate coefficient for events in which a node} \\
 &\quad \text{gains the trait through unfaithful horizontal transmission,} \\
 h'_- &= \text{rate coefficient for events in which a node} \\
 &\quad \text{loses the trait through unfaithful horizontal transmission.}
 \end{aligned} \tag{S1}$$

The exact mathematical interpretation of these rates is as follows:

We focus on a node which is connected to  $k$  other nodes,  $n(t) \leq k$  of which are in the opposite state to that of the focal node at time  $t$ .

If the focal node is the state +1, then the rate at which it flips to -1 is

$$T_{n,k,t}^- = \underbrace{h_- \frac{n(t)}{k}}_{\text{Correctly copy absence from a neighbour}} + \underbrace{h'_- \frac{k-n(t)}{k}}_{\text{Incorrectly copy presence as absence}} + \underbrace{v'_-}_{\text{Spontaneously lose existing trait}}, \quad (\text{S2})$$

where  $n(t)$  is the number of neighbours of the focal node in the down-state.

If the focal node is in state -1 then the rate at which nodes flip to state +1 is

$$T_{n,k,t}^+ = \underbrace{h_+ \frac{n(t)}{k}}_{\text{Correctly copy presence from a neighbour}} + \underbrace{h'_+ \frac{k-n(t)}{k}}_{\text{Incorrectly copy absence as presence}} + \underbrace{v'_+}_{\text{Spontaneously acquire the trait}}, \quad (\text{S3})$$

noting that  $n(t)$  is now the number of neighbours of the focal node in the up-state.

Fig. 1 in the main paper shows a pictorial representation of these dynamics.

## B. Notation

We will see that the six model parameters in Eq. (S1) ultimately only appear in specific combinations. Here we define shorthands for these combinations in order to reduce the notation later on,

$$\begin{aligned} \alpha &= (h'_+ - h'_-) + (v'_+ - v'_-), \\ \beta &= (h'_+ + h'_-) + (v'_+ + v'_-), \\ \gamma_- &= h_- - h'_-, \\ \gamma_+ &= h_+ - h'_+, \\ \gamma &= \frac{1}{2}(\gamma_+ + \gamma_-), \\ \delta &= h'_+ + v'_+, \\ \kappa &= h'_- + v'_-, \\ \eta &= (v'_+ + v'_-) + \frac{1}{2}[(h_+ + h_-) + (h'_+ + h'_-)]. \end{aligned} \quad (\text{S4})$$

Throughout most of the mathematical analysis we will make the restriction

$$\gamma_+ = \gamma_-. \quad (\text{S5})$$

This is required to proceed analytically. The restriction means that we assume

$$h_+ + h'_- = h_- + h'_+. \quad (\text{S6})$$

The coefficients on the left are those for faithful and unfaithful copying from a node in the +1 state. The coefficients on the right are those for events in which faithful or unfaithful copying occurs from a node in state -1. Broadly speaking, the constraint means that in the horizontal processes, the total rate of copying from +1 nodes is equal to the total rate of copying from -1 nodes.

## C. Relation of the present setup to the model in [S1]

In [S1] time is discrete and the model parameters are  $q, p_I, p_E, p'_I$  and  $p'_E$ . More precisely a spin is chosen at random from all spins (with equal probability). Then a vertical event is attempted with probability  $1 - q$ , or a horizontal event with probability  $q$ .

If a vertical event occurs and the spin is in state -1, then it flips to +1 with probability  $p_I$ . If the spin is in state +1, then it flips to -1 with probability  $p_E$  in the vertical event.

In a horizontal event the focal spin attempts to copy the state of a random nearest neighbour. If that neighbour is in state  $-1$ , then a transmission error occurs with probability  $p'_I$ , the focal spin then takes value  $+1$ . If the neighbour is in state  $+1$ , then a copying error occurs with probability  $p'_E$  and the focal spin takes value  $+1$ . Otherwise the horizontal copying is faithful.

This leads to transition probabilities (once the spin to be updated has been picked) with analogous structure to Eqs. (S2) and (S3):

$$T_{n,k,t}^- = q(1 - p'_I) \frac{n(t)}{k} + qp'_E \frac{k - n(t)}{k} + (1 - q)p_E, \quad (\text{S7a})$$

$$T_{n,k,t}^+ = q(1 - p'_E) \frac{n(t)}{k} + qp'_I \frac{k - n(t)}{k} + (1 - q)p_I, \quad (\text{S7b})$$

where  $n(t)$  is the number of neighbours in the opposite state to the focal node at time  $t$ .

Comparing Eqs. (S7a) and (S7b) with (S2) and (S3), we find the following mapping between the two parameter sets:

$$\begin{aligned} (1 - q)p_I &\leftrightarrow v'_+, \\ (1 - q)p_E &\leftrightarrow v'_-, \\ qp'_I &\leftrightarrow h'_+, \\ qp'_E &\leftrightarrow h'_-, \\ q(1 - p'_I) &\leftrightarrow h_-, \\ q(1 - p'_E) &\leftrightarrow h_+. \end{aligned} \quad (\text{S8})$$

Is it clear that a re-scaling of all rates  $v'_\pm, h_\pm, h'_\pm$  by a common positive factor only changes the units of time, but not the actual model dynamics. Therefore, we can re-write Eq. (S8) as

$$\begin{aligned} (1 - q)p_I &= \chi v'_+, \\ (1 - q)p_E &= \chi v'_-, \\ qp'_I &= \chi h'_+, \\ qp'_E &= \chi h'_-, \\ q(1 - p'_I) &= \chi h_-, \\ q(1 - p'_E) &= \chi h_+, \end{aligned} \quad (\text{S9})$$

with  $\chi > 0$ .

We note that the constraint  $h_+ + h'_- = h_- + h'_+$  in Eq. (S5) is automatically fulfilled by the setup in [S1], as both sides add up to  $q/\chi$  using the replacements in Eq. (S9).

Assume now we are given non-negative rates  $v'_\pm, h_\pm, h'_\pm$  such that  $h_+ + h'_- = h_- + h'_+$ . We now show that we can always find  $p_E, p_I, p'_E, p'_I, q \in [0, 1]$  and  $\chi > 0$  such that Eq. (S9) holds.

To do this, define

$$\Sigma \equiv h_+ + h'_- = h_- + h'_+. \quad (\text{S10})$$

Then set

$$p'_I = \frac{h'_+}{\Sigma}, \quad (\text{S11a})$$

$$p'_E = \frac{h'_-}{\Sigma}. \quad (\text{S11b})$$

Because of Eq. (S10) these quantities both take values between zero and one.

We also choose  $q$  as follows:

$$q = \chi \Sigma. \quad (\text{S12})$$

The factor  $\chi$  is not determined at this point, but has to be chosen such that  $\chi < 1/\Sigma$  (to ensure  $q < 1$ ).

One can directly check that the combination of Eqs. (S11a), (S11b) and (S12) ensures that the last four relations in Eqs. (S9) are fulfilled.

We now further fix  $p_I$  and  $p_E$  as follows:

$$p_I = \frac{\chi}{1 - \chi\Sigma} v'_+, \quad (\text{S13a})$$

$$p_E = \frac{\chi}{1 - \chi\Sigma} v'_-. \quad (\text{S13b})$$

Using  $q = \chi\Sigma$  [Eq. (S10)], this ensures that the first two relations in Eqs. (S9) are fulfilled. For given  $\Sigma$  and  $v'_\pm$  we can always choose  $\chi > 0$  small enough so that  $p_I$  and  $p_E$  in Eqs. (S13a) and (S13b) are less than one.

## S2. Complete network

We consider the model on a complete network (all-to-all interaction) with  $N$  nodes. There is then no notion of space, and any node is the nearest neighbour of any other node. Each node  $i$  can be in one of two states  $s_i \in \{-1, 1\}$ , and at each time the overall configuration of the model is given by  $\mathbf{s} = (s_1, \dots, s_N)$ . The system only has one degree of freedom, the number of nodes  $n$  in state +1,

$$n = \sum_i \frac{1 + s_i}{2}, \quad (\text{S14})$$

where we sum over all nodes  $i$  in the network.

### A. Transition rates in continuous-time and master equation

The dynamics proceeds through transitions  $n \rightarrow n \pm 1$ , with the following rates:

$$\begin{aligned} T_n^+ &= \underbrace{h_+ \frac{n(N-n)}{N}}_{\text{Correctly copy presence}} + \underbrace{h'_+ \frac{(N-n)^2}{N}}_{\text{Incorrectly copy absence as presence}} + \underbrace{v'_+ (N-n)}_{\text{Spontaneously acquire the trait}} \\ &= \delta N + (\gamma_+ - \delta)n - \gamma_+ \frac{n^2}{N}, \end{aligned} \quad (\text{S15a})$$

$$\begin{aligned} T_n^- &= \underbrace{h_- \frac{n(N-n)}{N}}_{\text{Correctly copy absence}} + \underbrace{h'_- \frac{n^2}{N}}_{\text{Incorrectly copy presence as absence}} + \underbrace{v'_- n}_{\text{Spontaneously lose an existing trait}} \\ &= (h_- + v'_-)n - \gamma_- \frac{n^2}{N}, \end{aligned} \quad (\text{S15b})$$

where we have used the expressions in Eq. (S4).

Writing  $P(n, t)$  for the probability of finding the system in state  $n$  at time  $t$ , one has the master equation

$$\frac{d}{dt} P(n, t) = T_{n-1}^+ P(n-1, t) + T_{n+1}^- P(n+1, t) - [T_n^+ + T_n^-] P(n, t). \quad (\text{S16})$$

### B. Magnetisation

We define the magnetisation as

$$m = \frac{1}{N} \sum_i s_i. \quad (\text{S17})$$



Using the master equation from Eq. (S16) we can derive a differential equation for the average magnetisation  $\langle m(t) \rangle$ , where  $\langle \dots \rangle$  represents an average over  $P(n, t)$ , i.e. realisations of the dynamics. To do this, we first see from Eqs. (S14) and (S17),

$$m = 2\frac{n}{N} - 1 \implies \frac{d\langle m(t) \rangle}{dt} = \frac{2}{N} \frac{d\langle n(t) \rangle}{dt}. \quad (\text{S18})$$

Then, from the master equation for the probability distribution  $P(n, t)$  [Eq. (S16)],

$$\begin{aligned} \frac{d\langle n(t) \rangle}{dt} &= \frac{d}{dt} \sum_n nP(n, t) = \sum_n n \frac{dP(n, t)}{dt} \\ &= \sum_n n \left\{ T_{n-1}^+ P(n-1, t) + T_{n+1}^- P(n+1, t) - [T_n^+ + T_n^-] P(n, t) \right\} \\ &= \sum_n \left\{ (n+1)T_n^+ P(n, t) + (n-1)T_n^- P(n, t) - [T_n^+ + T_n^-] P(n, t) \right\} \\ &= \sum_n [T_n^+ - T_n^-] P(n, t) \\ &= \langle T_{n(t)}^+ \rangle - \langle T_{n(t)}^- \rangle. \end{aligned} \quad (\text{S19})$$

Thus,

$$\frac{d\langle m(t) \rangle}{dt} = \frac{2}{N} \left[ \langle T_{n(t)}^+ \rangle - \langle T_{n(t)}^- \rangle \right]. \quad (\text{S20})$$

Assuming the population is of infinite size, we can ignore fluctuations, and make the following *deterministic approximation*:

$$\langle T_{n(t)}^\pm \rangle = T_{\langle n(t) \rangle}^\pm. \quad (\text{S21})$$

We then obtain the following rate equation for the magnetisation by substituting the explicit expressions for the rates from Eqs. (S15a) and (S15b) into Eq. (S20),

$$\frac{d\langle m(t) \rangle}{dt} = 2\delta + 2(\gamma_+ - \delta - h_- - v'_-) \frac{\langle n(t) \rangle}{N} - 2(\gamma_+ - \gamma_-) \frac{\langle n(t) \rangle^2}{N^2}. \quad (\text{S22})$$

We now make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , which closes the above equation and gives

$$\begin{aligned} \frac{d\langle m(t) \rangle}{dt} &= 2\delta - 2\beta \langle n(t) \rangle \\ &= 2\delta - 2\beta \frac{1 + \langle m(t) \rangle}{2} \\ &= \alpha - \beta \langle m(t) \rangle. \end{aligned} \quad (\text{S23})$$

Thus (assuming  $\beta \neq 0$ ) the average steady-state magnetisation is

$$\langle m \rangle_{\text{st}} = \frac{\alpha}{\beta} = \frac{(h'_+ - h'_-) + (v'_+ - v'_-)}{(h'_+ + h'_-) + (v'_+ + v'_-)}. \quad (\text{S24})$$

We note that the case  $\beta = 0$  corresponds to all unfaithful model parameters being 0. This is the standard voter model where the average steady-state magnetisation is simply the initial average magnetisation,  $\langle m \rangle_{\text{st}} = \langle m \rangle_0$ .

Equation (S23) can also be solved to give the full solution for  $m(t)$ ,

$$\langle m(t) \rangle = \langle m \rangle_{\text{st}} - (\langle m \rangle_{\text{st}} - \langle m \rangle_0) e^{-\beta t}. \quad (\text{S25})$$

### C. Interface density

We define interface density as the proportion of active links, i.e the proportion of links connecting two nodes in different states. For the complete network we have

$$\sigma = \frac{\# \text{ active links}}{\# \text{ links}} = \frac{n(N-n)}{\frac{N(N-1)}{2}} = \frac{1}{2}(1-m^2), \quad (\text{S26})$$

where  $m$  is given in Eq. (S18). The definition in Eq. (S26) holds for all values of  $N$ . For any time  $t$  we thus have, again for all  $N$ ,

$$\langle \sigma(t) \rangle = \frac{1}{2} (1 - \langle m(t)^2 \rangle) \quad (\text{S27})$$

Under the deterministic approximation, Eq. (S21), we have  $\langle m(t)^2 \rangle = \langle m(t) \rangle^2$ , thus

$$\langle \sigma(t) \rangle = \frac{1}{2} (1 - \langle m(t) \rangle^2), \quad (\text{S28})$$

with  $\langle m(t) \rangle$  as given in Eq. (S25). The average steady-state interface density for infinite complete networks is then given by

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} (1 - \langle m \rangle_{\text{st}}^2). \quad (\text{S29})$$

### S3. Pair approximation

Next, we analyse the model in the so-called *pair approximation*. This follows the lines of [S2]. In interacting-agent models of the type we are considering here the pair approximation usually captures the behaviour of the model to good accuracy on infinite uncorrelated networks [S5] (i.e. networks where there is no preference for nodes to attach to nodes with similar degree [S6]).

#### A. Transition rates for individual nodes

We start from the rates in Eqs. (S2) and (S3), for a focal node of degree  $k$  and with  $n(t)$  neighbours in the opposite state to that of the focal node at time  $t$ .

We can write

$$\begin{aligned} T_{n,k,t}^- &= h_- \frac{n(t)}{k} + h'_- \frac{k-n(t)}{k} + v'_- \\ &= \gamma_- \frac{n(t)}{k} + \kappa, \end{aligned} \quad (\text{S30})$$

and

$$\begin{aligned} T_{n,k,t}^+ &= h_+ \frac{n(t)}{k} + h'_+ \frac{k-n(t)}{k} + v'_+ \\ &= \gamma_+ \frac{n(t)}{k} + \delta. \end{aligned} \quad (\text{S31})$$

We have used the shorthands in Eq. (S4).

## B. Magnetisation

The change in the average magnetisation is described by the differential equation,

$$\frac{d\langle m(t) \rangle}{dt} = \sum_k \sum_{s=\pm} NP_k \langle \rho_s(t) \rangle \sum_{n=0}^k T_{n,k,t}^{-s} B(n|k, s, t) \Delta m(s). \quad (\text{S32})$$

We use  $P_k$  to denote the degree distribution of the network, with normalisation  $\sum_k P_k = 1$ . The quantity  $\langle \rho_s(t) \rangle$  is the average proportion of nodes in state  $s$ . Thus  $NP_k \langle \rho_s(t) \rangle$  is the expected number of nodes with degree  $k$  and in state  $s$  at time  $t$ . Further,  $B(n|k, s, t)$  is the probability that a node of degree  $k$  and in state  $s \in \{-1, 1\}$  has  $n$  neighbours in the opposite state at time  $t$ . Finally,  $\Delta m(s) = (-2s)/N$  is the amount by which the magnetisation changes if a node in state  $s$  flips to  $-s$ .

Substituting in the expression for  $\Delta m(s)$  and expanding the sum over  $s$ , Eq. (S32) becomes

$$\frac{d\langle m(t) \rangle}{dt} = \sum_k P_k \sum_{n=0}^k \left\{ \left[ 1 - \langle m(t) \rangle \right] T_{n,k,t}^+ B(n|k, -, t) - \left[ 1 + \langle m(t) \rangle \right] T_{n,k,t}^- B(n|k, +, t) \right\}. \quad (\text{S33})$$

The pair approximation consists of assuming that  $B(n|k, s, t)$  is a binomial distribution. This means we assume that the states of nearest neighbours of a focal node are uncorrelated. The single-event probability is  $P(-s|s, t)$ , which is the conditional probability that a neighbour of a node is in state  $-s$  given that the node is in state  $s$  at time  $t$ . We have,

$$B(n|k, s, t) = \binom{k}{n} P(-s|s, t)^n \left[ 1 - P(-s|s, t) \right]^{k-n}. \quad (\text{S34})$$

Writing  $\mu$  for the average degree of the network, the quantity  $P(-s|s, t)$  can be obtained as follows,

$$P(-s|s, t) = \frac{\langle \sigma(t) \rangle \cdot \frac{\mu N}{2}}{\langle \rho_s(t) \rangle \cdot \mu N} = \frac{\langle \sigma(t) \rangle}{2 \langle \rho_s(t) \rangle} = \frac{\langle \sigma(t) \rangle}{1 + s \langle m(t) \rangle}, \quad (\text{S35})$$

The numerator in Eq. (S35) is the average total number of active links in the network (on average there are  $\frac{\mu N}{2}$  links overall, and a fraction  $\langle \sigma(t) \rangle$  of these are active). The denominator ( $\langle \rho_s(t) \rangle \cdot \mu N$ ) is the average total number of links emanating from all nodes in state  $s$ .

The first and second moments of the binomial distribution in Eq. (S34) are then [S7]

$$E[n]_{k,s,t} = \frac{k \langle \sigma(t) \rangle}{1 + s \langle m(t) \rangle}, \quad (\text{S36a})$$

$$\begin{aligned} E[n^2]_{k,s,t} &= \frac{k \langle \sigma(t) \rangle}{1 + s \langle m(t) \rangle} \left( 1 - \frac{\langle \sigma(t) \rangle}{1 + s \langle m(t) \rangle} \right) + \frac{k^2 \langle \sigma(t) \rangle^2}{(1 + s \langle m(t) \rangle)^2} \\ &= \frac{k \langle \sigma(t) \rangle}{1 + s \langle m(t) \rangle} + \frac{k(k-1) \langle \sigma(t) \rangle^2}{(1 + s \langle m(t) \rangle)^2}. \end{aligned} \quad (\text{S36b})$$

Using these moments, and the rates from Eqs. (S30) and (S31), we can simplify Eq. (S33) to obtain

$$\begin{aligned} \frac{d\langle m(t) \rangle}{dt} &= \sum_k P_k \left\{ \left( 1 - \langle m(t) \rangle \right) \left[ \gamma_+ \frac{E[n]_{k,-,t}}{k} + \delta \right] - \left( 1 + \langle m(t) \rangle \right) \left[ \gamma_- \frac{E[n]_{k,+,t}}{k} + \kappa \right] \right\} \\ &= \alpha - \beta \langle m(t) \rangle + (\gamma_+ - \gamma_-) \langle \sigma(t) \rangle. \end{aligned} \quad (\text{S37})$$

We now again make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , after which Eq. (S37) is identical to Eq. (S23) for the complete network. Thus the solution for  $\langle m(t) \rangle$  within the pair approximation is as in Eq. (S25), and the steady-state value is given by Eq. (S24).

## C. Interface density

We consider again a node of degree  $k$  and with  $n(t)$  active links to its neighbours at time  $t$ , i.e.  $n(t)$  of its neighbours are of in the opposite state. If this focal node changes state, the change in the number of active

links is

$$\Delta n(t) = n(t+1) - n(t) = [k - n(t)] - n(t) = k - 2n(t), \quad (\text{S38})$$

Given that the total number of links in the network is  $\frac{\mu N}{2}$  on average (where we recall that  $\mu$  is the mean degree), the expected change in interface density associated with the flip of the focal node is

$$\Delta \sigma(n, k, t) = \frac{k - 2n(t)}{\frac{\mu N}{2}} = \frac{2[k - 2n(t)]}{\mu N}. \quad (\text{S39})$$

Similar to the magnetisation, the rate of change of the interface density can be written

$$\begin{aligned} \frac{d\langle \sigma(t) \rangle}{dt} &= \sum_k \sum_{s=\pm} NP_k \langle \rho_s(t) \rangle \sum_{n=0}^k T_{n,k,t}^{-s} B(n|k, s, t) \Delta \sigma(n, k, t) \\ &= \frac{1}{\mu} \sum_k P_k \sum_{n=0}^k \left\{ [1 - \langle m(t) \rangle] T_{n,k,t}^+ B(n|k, -, t) \right. \\ &\quad \left. + [1 + \langle m(t) \rangle] T_{n,k,t}^- B(n|k, +, t) \right\} [k - 2n(t)]. \end{aligned} \quad (\text{S40})$$

As in Sec. S3B we use Eqs. (S30), (S31), (S36a) and (S36b) to find

$$\begin{aligned} \frac{d\langle \sigma(t) \rangle}{dt} &= \beta - \alpha \langle m(t) \rangle + 2 \langle \sigma(t) \rangle \left[ \gamma \left( 1 - \frac{2}{\mu} \right) - \beta \right] \\ &\quad - 2 \left( 1 - \frac{1}{\mu} \right) \langle \sigma(t) \rangle^2 \left[ \frac{2\gamma + (\gamma_+ - \gamma_-) \langle m(t) \rangle}{1 - \langle m(t) \rangle^2} \right]. \end{aligned} \quad (\text{S41})$$

We now use the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , to write

$$\frac{d\langle \sigma(t) \rangle}{dt} = \beta - \alpha \langle m(t) \rangle + 2 \langle \sigma(t) \rangle \left[ \gamma \left( 1 - \frac{2}{\mu} \right) - \beta \right] - 4\gamma \left( 1 - \frac{1}{\mu} \right) \frac{\langle \sigma(t) \rangle^2}{1 - \langle m(t) \rangle^2}. \quad (\text{S42})$$

It is possible to solve this equation numerically [this is how the analytical trajectories of Fig. 2(a) in the main paper were obtained].

The steady-state solution can be found by equating the right-hand side of Eq. (S42) to zero, and, based on Eq. (2), substituting in the following ansatz,

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{PA}}(\tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right). \quad (\text{S43})$$

Here  $H_{\text{PA}}(\tau)$  is, at this stage, an unknown function of the network and the model parameters (the subscript ‘PA’ stands for ‘pair approximation’). This results in the following quadratic equation for  $H_{\text{PA}}(\tau)$ :

$$(\mu - 1)H_{\text{PA}}(\tau)^2 + [\mu\tau - (\mu - 2)]H_{\text{PA}}(\tau) - \mu\tau = 0, \quad (\text{S44})$$

which can be solved to give

$$H_{\text{PA}}(\tau) = \frac{\mu - 2 - \mu\tau \pm \sqrt{(\tau + 1)^2 \mu^2 - 4(\mu - 1)}}{2(\mu - 1)}, \quad (\text{S45})$$

where

$$\tau = \frac{\beta}{\gamma} = \frac{(h'_+ + h'_-) + (v'_+ + v'_-)}{\frac{1}{2} [(h_+ - h'_+) + (h_- - h'_-)]}. \quad (\text{S46})$$

We refer to  $\tau$  as the *noise-to-order* ratio, see Sec. S8 for a more detailed discussion of the general interpretation of  $\tau$  and the order parameter  $H(\tau)$ .

## D. Further discussion of the function $H_{\text{PA}}(\tau)$ obtained from the pair approximation

### 1. Regions of validity of the two branches of the solution

We note that, Eq. (S45) has both a positive and negative branch for a given value of  $\mu$ . We denote these  $H_{\text{PA}}^{\pm}(\tau)$ . Given that  $H(\tau)$  must be real we require the object under the square root in Eq. (S45) to be non-negative. This means that we must either have  $\tau \leq -1 - \frac{2}{\mu}\sqrt{\mu-1}$  or  $\tau \geq -1 + \frac{2}{\mu}\sqrt{\mu-1}$ . An interpretation of  $\tau$ , and a discussion of the range of values this parameter can take in our model can be found in Sec. S8B. This includes the possibility of negative values of  $\tau$ .

In the limit  $\tau \rightarrow -\infty$  we find  $H_{\text{PA}}^+(\tau) \rightarrow \infty$ , whereas  $H_{\text{PA}}^-(\tau) \rightarrow 1$ . In simulations we observe  $H(\tau) \rightarrow 1$  as  $\tau \rightarrow -\infty$ . Therefore we use  $H_{\text{PA}}^-(\tau)$  in the region  $\tau \leq -1 - \frac{2}{\mu}\sqrt{\mu-1}$ .

In the limit  $\tau \rightarrow \infty$  we find  $H_{\text{PA}}^-(\tau) \rightarrow -\infty$ , whereas  $H_{\text{PA}}^+(\tau) \rightarrow 1$ . In simulations we observe  $H(\tau) \rightarrow 1$  as  $\tau \rightarrow \infty$ . Therefore, we use  $H_{\text{PA}}^+(\tau)$  in the region  $\tau \geq -1 + \frac{2}{\mu}\sqrt{\mu-1}$ .

In Fig. S1 we show both branches of  $H_{\text{PA}}(\tau)$  for different values of  $\mu$ .

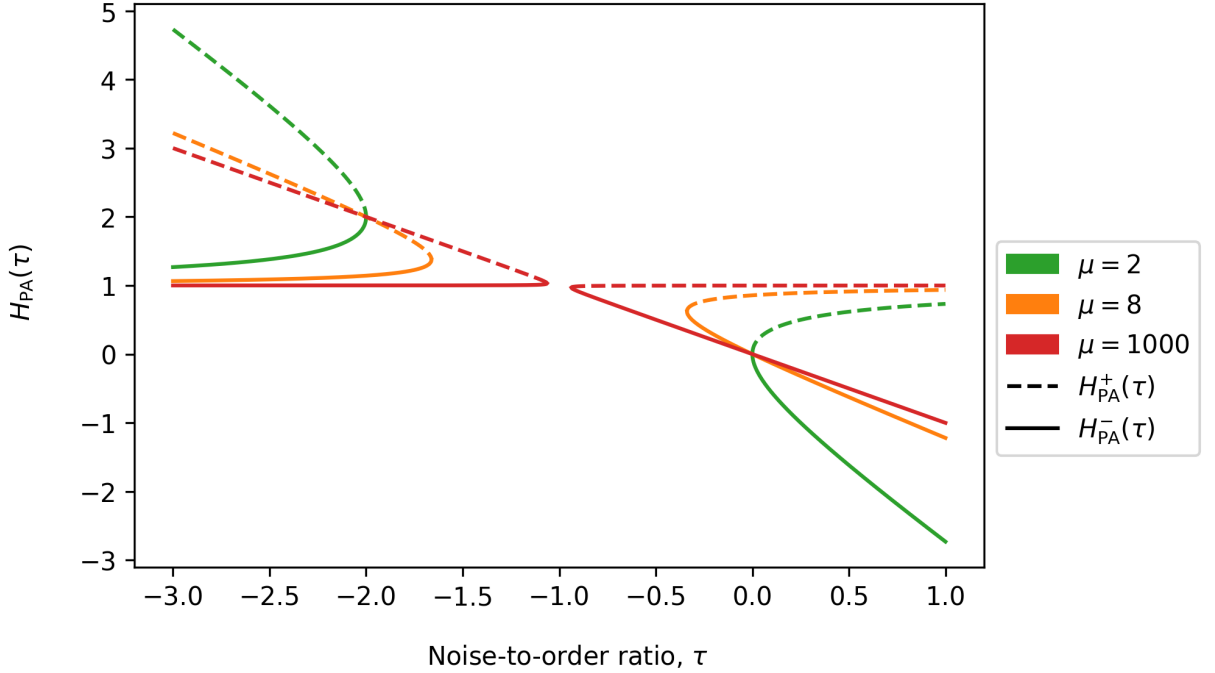


FIG. S1. Plots of  $H_{\text{PA}}^{\pm}(\tau)$  from Eq. (S45) as a function of the noise-to-order ratio,  $\tau$ , for different values of  $\mu$ .  $H_{\text{PA}}(\tau)$  only takes real values in the regions  $\tau \leq -1 - \frac{2}{\mu}\sqrt{\mu-1}$  or  $\tau \geq -1 + \frac{2}{\mu}\sqrt{\mu-1}$ . Solid lines are  $H_{\text{PA}}^-(\tau)$  (applicable for  $\tau \leq -1 - \frac{2}{\mu}\sqrt{\mu-1}$ ), dashed lines are  $H_{\text{PA}}^+(\tau)$  (applicable for  $\tau \geq -1 + \frac{2}{\mu}\sqrt{\mu-1}$ ).

### 2. Comparison against simulations

We are principally interested in  $\tau \geq 0$ , so we do not consider the negative branch of Eq. (S45) further. In Fig. S2 we demonstrate how the positive branch of Eq. (S45) compares to simulation for different size degree-regular (left), Erdős–Rényi (middle), and Barabási–Albert (right) networks.

As  $\tau \rightarrow \infty$  we see that both  $H(\tau) \rightarrow 1$  both in simulations and that  $H_{\text{PA}}^+(\tau) \rightarrow 1$ . For sufficiently low values of  $\tau$ , simulation results deviate from the theory. This is because simulations are performed on finite size networks and/or because of inaccuracies in the pair approximation.

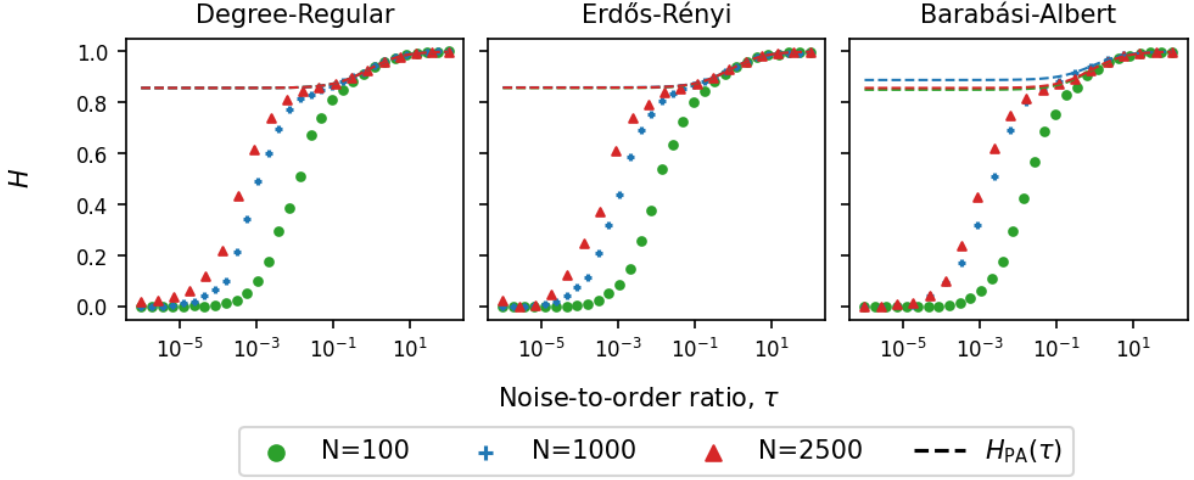


FIG. S2. Plots of  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)]$ , which characterises the scattering, against the noise-to-order ratio  $\tau$  for networks of size  $N$ . Degree-regular with  $\mu = 8$  (left), Erdős-Rényi with  $p = 8/N$  (middle), Barabási-Albert with 4 new links established upon adding a new node (right). For all three types of network we have  $\mu \approx 8$  (deviations are due to the finite size of the networks). The dashed lines are the results  $H_{\text{PA}}^+(\tau)$  from the pair approximation, Eq. (S45). Note that there are 3 lines on each plot, as varying  $N$  can change  $\mu$  which results in a different  $H_{\text{PA}}(\tau)$ . For degree-regular networks,  $\mu = 8$  exactly always so the lines are identical, however for Barabási-Albert networks the difference is noticeable. Each marker is the result from averaging 1,000 independent simulations in the steady-state at a particular  $\tau$  value. Model parameters for a particular  $\tau$  are generated via the algorithm detailed in Sec. S10 B.

As  $\tau \rightarrow 0$  we observe in simulations on finite networks that  $H(\tau) \rightarrow 0$ . However  $H_{\text{PA}}(\tau) \rightarrow \frac{\mu-2}{\mu-1}$  as  $\tau \rightarrow 0$ . We note that this is equivalent to the long-lived plateau of the interface density reported in [S2]. We find that the agreement between simulation results and  $H_{\text{PA}}(\tau)$  improves with increasing system size  $N$  as expected.

### 3. Limits on the inference of $\tau$ from $H(\tau)$

It is not always possible to map a value of  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)]$  from simulations (or data) to a value of  $\tau$  within the pair approximation. For example, for  $\mu = 8$  no value of  $\tau$  is associated with values of  $H(\tau)$  between zero and approximately 0.85 [the limiting value of  $H_{\text{PA}}(\tau)$  for  $\tau \rightarrow 0$  in Fig. S2]. The fact that not all values of  $H(\tau)$  are reached by the applicable branches of  $H_{\text{PA}}(\tau)$  can also be seen in Fig. S1.

### E. Limit of all-to-all interaction

As a sanity check we verify that the results in the pair approximation reduce to those for the complete network in the limit  $\mu \rightarrow \infty$ . Taking this limit in Eq. (S42) we find

$$\frac{d \langle \sigma(t) \rangle}{dt} = \beta - \alpha m(t) + 2\sigma(\gamma - \beta) - 4\gamma \frac{\langle \sigma(t) \rangle^2}{1 - \langle m(t) \rangle^2}. \quad (\text{S47})$$

We now verify that the result in Eq. (S28) is a solution of Eq (S47). Substituting Eq. (S28) into the right-hand side of Eq. (S47) gives

$$\frac{d \langle \sigma(t) \rangle}{dt} = -[\alpha - \beta \langle m(t) \rangle] \langle m(t) \rangle. \quad (\text{S48})$$

On the other hand, differentiating the solution in Eq. (S28) with respect to time gives

$$\frac{d \langle \sigma(t) \rangle}{dt} = -\langle m(t) \rangle \frac{d \langle m(t) \rangle}{dt}. \quad (\text{S49})$$



Comparing Eqs. (S48) and (S49) it must then be true that

$$\frac{d\langle m(t) \rangle}{dt} = \alpha - \beta \langle m(t) \rangle. \quad (\text{S50})$$

This is the same differential equation we obtained on the complete network, i.e. Eq. (S23), and we have already established in Eq. (S37) that  $\langle m(t) \rangle$  in the pair approximation is the same as that on the complete network. Therefore we have shown that the result for the interface density from the pair approximation reduces to that for the complete network in the limit  $\mu \rightarrow \infty$ .

We can also take the limit  $\mu \rightarrow \infty$  of  $H_{\text{PA}}(\tau)$  from Eq. (S45), and we find  $H_{\text{PA}}(\tau) \rightarrow 1$  for all  $\tau$ , i.e. we obtain Eq. (S29).

#### S4. Annealed approximation

We now analyse the model in the so-called *annealed approximation*, which works well for large but finite-size uncorrelated networks. We broadly follow the steps of [S4], but we note a difference in notation. While this earlier reference uses  $s_i \in \{0, 1\}$  for the binary states of the nodes, we use  $s_i \in \{-1, 1\}$ , consistent with the rest of our paper.

##### A. Nature of the approximation and the rate equations

The annealed approximation centers around replacing an adjacency matrix  $\mathbf{A}$  (representing a network with a given degree sequence) with a weighted adjacency matrix  $\tilde{\mathbf{A}}$ . The elements  $\tilde{A}_{ij}$  are the probabilities that nodes  $i$  and  $j$  are connected over an ensemble of networks generated from the configuration model [S8] using the degree sequence of the original network.

The elements of the matrix  $\tilde{\mathbf{A}}$  are the probabilities that node  $i$  and  $j$  are connected in the ensemble of networks generated by the configuration model, i.e. [S9, (13.32)]

$$\tilde{A}_{ij} \approx \frac{k_i k_j}{N\mu} \quad (\text{S51})$$

With this we can replace any summation over nearest neighbours of a node as follows (we write  $j \in N(i)$  if  $j$  is a nearest neighbour of  $i$ ):

$$\sum_{j \in N(i)} s_j \rightarrow \frac{k_i}{\mu N} \sum_j k_j s_j. \quad (\text{S52})$$

We will refer to this as the *annealed approximation*. The approximation conserves the initial degree sequence,

$$k_i = \sum_{j=1}^N \tilde{A}_{ij} = k_i \frac{1}{\mu} \left( \frac{1}{N} \sum_{j=1}^N k_j \right) = k_i, \quad (\text{S53})$$

and thus also the total number of links.

We can now derive the rates at which nodes change state using the annealed approximation from Eq. (S52). To ease the notation we write the final rates using the expressions from Eq. (S4). If node  $i$  is in state  $s_i = -1$ , then it will flip to state  $s_i = +1$  with rate

$$\begin{aligned} T_i^+ &= v'_+ + \frac{h_+}{k_i} \sum_{j \in N(i)} \frac{1+s_j}{2} + \frac{h'_+}{k_i} \sum_{j \in N(i)} \frac{1-s_j}{2} \\ &= \delta + \frac{\gamma_+}{2} + \frac{\gamma_+}{2\mu N} \sum_m k_m s_m. \end{aligned} \quad (\text{S54})$$

We note that the first line of Eq. (S54) consists of three terms. The first is the rate coefficient at which spontaneous flipping from  $-1$  to  $+1$  occurs. The second term is the probability that a randomly chosen neighbour has state  $+1$ , multiplied by the rate coefficient for a  $-1$  node faithfully copying the state of a  $+1$  node. The final term is the probability that a randomly chosen neighbour has the state  $-1$ , multiplied by the rate coefficient for a  $-1$  node to erroneously copy the state of a  $+1$  node,  $h'_+$ . If the node is in state  $s_i = +1$  it will flip to  $s_i = -1$  with rate

$$\begin{aligned} T_i^- &= v'_- + \frac{h_-}{k_i} \sum_{j \in N(i)} \frac{1 + s_j}{2} + \frac{h'_-}{k_i} \sum_{j \in i} \frac{1 - s_j}{2} \\ &= \kappa + \frac{\gamma_-}{2} - \frac{\gamma_-}{2\mu N} \sum_m k_m s_m. \end{aligned} \quad (\text{S55})$$

With these definitions we then have the overall flip rate for spin  $i$  which depends on the spin configuration  $\mathbf{s} = (s_1, \dots, s_N)$ ,

$$W_i(\mathbf{s}) \equiv \frac{1 + s_i}{2} T_i^- + \frac{1 - s_i}{2} T_i^+. \quad (\text{S56})$$

### B. Master equation

To formulate the master equation for the model, it is useful to define the flip operators  $F_i$  acting on functions of  $\mathbf{s}$  by flipping the  $i^{\text{th}}$  spin,

$$F_i f(s_1, \dots, s_i, \dots, s_N) \equiv f(s_1, \dots, -s_i, \dots, s_N). \quad (\text{S57})$$

The master equation for  $P(\mathbf{s})$ , i.e. the probability that  $N$  nodes have a particular spin configuration, is then given by

$$\frac{dP(\mathbf{s})}{dt} = \sum_{i=1}^N \left\{ F_i \left[ W_i(\mathbf{s}) P(\mathbf{s}) \right] - W_i(\mathbf{s}) P(\mathbf{s}) \right\}. \quad (\text{S58})$$

The second term in this equation is the rate at which  $P(\mathbf{s})$  decreases. For this to happen the system must be in state  $\mathbf{s}$ , which happens with probability  $P(\mathbf{s})$ . Then any of the  $i$  spins must flip so that the state is no longer  $\mathbf{s}$ , which happens with rate  $W_i(\mathbf{s})$ . The first term gives the rate at which  $P(\mathbf{s})$  increases. For this to happen the system must be in a state where after flipping a single spin  $i$  it will be in state  $\mathbf{s}$ . The probability of being in such a state is  $F_i[P(\mathbf{s})]$ . The spin  $i$  must then flip, this happens with rate  $F_i[W_i(\mathbf{s})]$ . The product of these things then form the first term. Finally, we sum over all spins  $i$  that can flip in a single time-step.

### C. Magnetisation

Following [S4] we first introduce the following summation notation:  $\sum_{\mathbf{s}}$  refers to a sum over all possible spin combinations, i.e

$$\sum_{\mathbf{s}} = \sum_{s_1} \sum_{s_2} \dots \sum_{s_N}, \quad (\text{S59})$$

and  $\sum_{\mathbf{s}_j}$  refers to a sum over all possible spin combinations excluding node  $j$ ,

$$\sum_{\mathbf{s}_j} \equiv \sum_{s_1} \dots \sum_{s_{j-1}} \sum_{s_{j+1}} \dots \sum_{s_N}. \quad (\text{S60})$$

These two summation definitions are related by

$$\sum_{\mathbf{s}} = \sum_{\mathbf{s}_j} \sum_{s_j}. \quad (\text{S61})$$

We denote the average over the spin configuration  $\mathbf{s}$  as

$$\langle f(\mathbf{s}) \rangle = \sum_{\mathbf{s}} f(\mathbf{s}) P(\mathbf{s}). \quad (\text{S62})$$

Using the master equation, Eq. (S58), we have

$$\begin{aligned} \frac{d\langle s_i \rangle}{dt} &= \sum_{\mathbf{s}} s_i \frac{dP(\mathbf{s})}{dt} \\ &= \sum_{\mathbf{s}} s_i \sum_{j=1}^N \left\{ F_j[W_j P] - W_j P \right\}, \end{aligned} \quad (\text{S63})$$

where we omit the argument  $\mathbf{s}$  of the  $W_j$  and  $P$  on the right for simplicity. We now split the sum over  $j$  into summations of  $j = i$  and  $j \neq i$ ,

$$\begin{aligned} \frac{d\langle s_i \rangle}{dt} &= \sum_{\mathbf{s}} s_i \left\{ \left( F_i[W_i P] - W_i P \right) + \sum_{j \neq i}^N \left( F_j[W_j P] - W_j P \right) \right\} \\ &= \sum_{\mathbf{s}_i} \sum_{\mathbf{s}_i} s_i \left( F_i[W_i P] - W_i P \right) + \sum_{j \neq i}^N \sum_{\mathbf{s}_j} s_i \sum_{\mathbf{s}_j} \left( F_j[W_j P] - W_j P \right) \\ &= \sum_{\mathbf{s}} \left( -2s_i W_i P \right), \end{aligned} \quad (\text{S64})$$

where in the second line, the second term vanishes after performing the  $s_j$  summation, and the first term simplifies after the  $s_i$  sum. Evaluating the sum with Eq. (S62) we have

$$\frac{d\langle s_i \rangle}{dt} = -2 \langle W_i s_i \rangle. \quad (\text{S65})$$

This can also be understood intuitively. The quantity  $W_i$  on the right is the probability with which spin  $i$  flips from  $s_i$  to  $-s_i$ . The change of  $s_i$  in such an event is  $-2s_i$ .

Using the definition of  $W_i$  in Eq. (S56), we then have

$$\frac{d\langle s_i \rangle}{dt} = \langle T_i^+ - T_i^- \rangle - \langle (T_i^+ + T_i^-) s_i \rangle. \quad (\text{S66})$$

We now use the explicit expressions for the rates defined in Eqs. (S54) and (S55) to write

$$\frac{d\langle s_i \rangle}{dt} = \alpha + \frac{1}{2}(\gamma_+ - \gamma_-) - (\beta + \gamma) \langle s_i \rangle + \frac{\gamma}{\mu N} \sum_m k_m \langle s_m \rangle - \frac{(\gamma_+ + \gamma_-)}{2\mu N} \sum_m k_m \langle s_i s_m \rangle. \quad (\text{S67})$$

Then we make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , and find

$$\frac{d\langle s_i \rangle}{dt} = \alpha - (\beta + \gamma) \langle s_i \rangle + \frac{\gamma}{\mu N} \sum_m k_m \langle s_m \rangle. \quad (\text{S68})$$

In the steady-state, this gives the following set of  $N$  simultaneous equations:

$$\alpha - (\beta + \gamma) \langle s_i \rangle + \frac{\gamma}{\mu N} \sum_m k_m \langle s_m \rangle = 0. \quad (\text{S69})$$

One can directly verify that

$$\langle s_i \rangle_{\text{st}} = \frac{-\alpha}{-(\gamma + \beta) + \sum_{i=1}^N A_i} = \frac{\alpha}{\beta} \quad (\text{S70})$$

for all  $i$  is a solution. This then also means

$$\langle m \rangle_{\text{st}} = \frac{1}{N} \sum_{i=1}^N \langle s_i \rangle_{\text{st}} = \frac{\alpha}{\beta}. \quad (\text{S71})$$

We now show that this is also the only solution of Eq. (S69). If we define

$$\mathcal{A}_i \equiv \frac{\gamma k_i}{\mu N}, \quad (\text{S72a})$$

$$\mathcal{B}_i \equiv \mathcal{A}_i - (\gamma + \beta), \quad (\text{S72b})$$

then we can cast Eq. (S69) as a matrix equation,

$$\begin{bmatrix} \mathcal{B}_1 & \mathcal{A}_2 & \cdots & \mathcal{A}_N \\ \mathcal{A}_1 & \mathcal{B}_2 & \cdots & \mathcal{A}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_1 & \mathcal{A}_2 & \cdots & \mathcal{B}_N \end{bmatrix} \begin{bmatrix} \langle s_1 \rangle_{\text{st}} \\ \langle s_2 \rangle_{\text{st}} \\ \vdots \\ \langle s_N \rangle_{\text{st}} \end{bmatrix} = \begin{bmatrix} -\alpha \\ -\alpha \\ \vdots \\ -\alpha \end{bmatrix}. \quad (\text{S73})$$

To prove the uniqueness of the above solution, we need to show that the matrix on the left has full rank, i.e. that its columns are linearly independent. We show this by contradiction.

Assume there are multiple solutions, i.e., there are coefficients  $c_1, \dots, c_N$  (which are not all equal to zero) such that

$$\begin{bmatrix} \mathcal{B}_1 & \mathcal{A}_2 & \cdots & \mathcal{A}_N \\ \mathcal{A}_1 & \mathcal{B}_2 & \cdots & \mathcal{A}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_1 & \mathcal{A}_2 & \cdots & \mathcal{B}_N \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (\text{S74})$$

Looking at the  $i^{\text{th}}$  and  $j^{\text{th}}$  row together, one concludes

$$c_i(\mathcal{B}_i - \mathcal{A}_i) = c_j(\mathcal{B}_j - \mathcal{A}_j) \quad (\text{S75})$$

for all pairs  $i, j$ . Given that  $\mathcal{B}_i - \mathcal{A}_i = -(\gamma + \beta)$  for all  $i$ , this then implies that all  $c_i$  must take the same value (we discard the special case  $\gamma + \beta = 0$ ). This common value can be non-zero only if the sum of elements in each row of the matrix in Eq. (S74) sum to zero, i.e. if

$$\begin{aligned} \mathcal{B}_i + \sum_{j \neq i}^N \mathcal{A}_j &= 0 \\ \implies -(\gamma + \beta) + \sum_{i=1}^N \mathcal{A}_i &= 0 \\ \implies -(\gamma + \beta) + \frac{\gamma}{\mu N} \sum_{i=1}^N k_i &= 0 \\ \implies \beta &= 0, \end{aligned} \quad (\text{S76})$$

where we have used the definitions in Eqs. (S72a) and (S72b). Therefore we are left with the condition  $\beta = 0$ . Keeping in mind the definition of  $\beta$  in Eq. (S4) this corresponds to a ‘pathological’ scenario, in which there is no spontaneous spin flipping (no vertical dynamics), and in which horizontal copying is without error. This is the standard voter model, which, for finite systems, will always end up in a consensus state. We note that  $\alpha = 0$  in this case as well.

In summary, for all cases of interest ( $\beta \neq 0$ ), the vector  $(c_1, \dots, c_N)$  is non-zero. Thus Eq. (S71) is the only non-trivial solution to Eq. (S69).

### D. Interface density

The average interface density can be expressed in the following form

$$\langle \sigma \rangle = \frac{\frac{1}{2} \sum_{ij} A_{ij} \left[ \frac{1+\langle s_i \rangle}{2} \frac{1-\langle s_j \rangle}{2} + \frac{1-\langle s_i \rangle}{2} \frac{1+\langle s_j \rangle}{2} \right]}{\frac{1}{2} \sum_{ij} A_{ij}}, \quad (\text{S77})$$

where the denominator is the total number of links and the numerator is the number of links connecting opposite spins. Using the annealed approximation, Eq. (S51), this can be written as

$$\langle \sigma \rangle = \frac{1}{2(\mu N)^2} \sum_{ij} k_i k_j [1 - \langle s_i s_j \rangle]. \quad (\text{S78})$$

In the steady-state we therefore have

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2(\mu N)^2} \sum_{ij} k_i k_j [1 - \langle s_i s_j \rangle_{\text{st}}]. \quad (\text{S79})$$

We now define the correlation matrix  $\Upsilon_{ij}$ ,

$$\Upsilon_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle, \quad (\text{S80})$$

which in the steady-state, since  $\langle s_i \rangle_{\text{st}} = \langle m \rangle_{\text{st}}$  for all  $i$  [see Eq. (S71)], can be written as

$$(\Upsilon_{ij})_{\text{st}} = \langle s_i s_j \rangle_{\text{st}} - \langle m \rangle_{\text{st}}^2. \quad (\text{S81})$$

Eq. (S79) then becomes

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} \left( 1 - \langle m \rangle_{\text{st}}^2 \right) - \frac{1}{2(\mu N)^2} \sum_{ij} k_i k_j (\Upsilon_{ij})_{\text{st}}. \quad (\text{S82})$$

The first term in this expression is the density of active links if the states of neighbouring spins were entirely uncorrelated, which is the same as the formula for the complete network [see Eq. (S29)]. The second term represents the reduction in interface density due to correlations introduced by the properties of the network. These appear as terms proportional to  $k_i k_j$  from the weighted adjacency matrix, Eq. (S51).

To find  $(\Upsilon_{ij})_{\text{st}}$ , we first formulate a differential equation for  $\langle s_i s_j \rangle$ . This is done in a similar manner to Sec. S4C. We use the master equation, Eq. (S58), to write

$$\begin{aligned} \frac{d \langle s_i s_j \rangle}{dt} &= \sum_{\mathbf{s}} s_i s_j \frac{dP(\mathbf{s})}{dt} \\ &= \sum_{\mathbf{s}} s_i s_j \sum_{k=1}^N \left\{ F_k[W_k P] - W_k P \right\}. \end{aligned} \quad (\text{S83})$$

Again, we have dropped the explicit dependence of  $W_k$  and  $P$  on  $\mathbf{s}$ . We first consider the term  $k \neq i, k \neq j$ . One has

$$\sum_{s_k} \left( F_k[W_k P] - W_k P \right) = 0, \quad (\text{S84})$$

for any fixed  $s_1, \dots, s_{k-1}, s_{k+1}, \dots, s_N$ . Thus this term does not contribute. If  $k = i \neq j$  we have

$$\sum_{s_i} s_j \left\{ \sum_{s_i} s_i \left( F_i[W_i P] - W_i P \right) \right\} = -2 \sum_{\mathbf{s}} s_i s_j W_i P = -2 \langle s_i s_j W_i \rangle. \quad (\text{S85})$$

Analogously the case  $k = j \neq i$  gives  $-2 \langle s_i s_j W_j \rangle$ . Finally for  $k = i = j$ , we note  $s_i^2 = 1$ , and therefore

$$\frac{d \langle s_i s_i \rangle}{dt} = 0. \quad (\text{S86})$$

Combining all of this, and substituting the expression for  $W_i$  from Eq. (S56), gives

$$\frac{d\langle s_i s_j \rangle}{dt} = \begin{cases} \langle (T_i^+ - T_i^-) s_j \rangle + \langle (T_j^+ - T_j^-) s_i \rangle - \langle q_{ij} s_i s_j \rangle & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}, \quad (\text{S87})$$

where  $q_{ij} = T_i^+ + T_i^- + T_j^+ + T_j^-$ . Using the Kronecker delta, Eq. (S87) can be written

$$\begin{aligned} \frac{d\langle s_i s_j \rangle}{dt} &= \langle (T_i^+ - T_i^-) s_j \rangle + \langle (T_j^+ - T_j^-) s_i \rangle \\ &\quad - \langle q_{ij} s_i s_j \rangle + 2\delta_{ij} \left[ \langle T_i^+ + T_i^- \rangle - \langle (T_i^+ - T_i^-) s_i \rangle \right]. \end{aligned} \quad (\text{S88})$$

We now substitute the explicit expressions for the rates, Eqs. (S54) and (S55), into Eq. (S88) which gives

$$\begin{aligned} \frac{d\langle s_i s_j \rangle}{dt} &= \left[ \alpha + \frac{1}{2}(\gamma_+ - \gamma_-) \right] (\langle s_i \rangle + \langle s_j \rangle) + \frac{\gamma}{\mu N} \sum_m k_m \left[ \langle s_m s_j \rangle + \langle s_m s_i \rangle \right] \\ &\quad - 2(\beta + \gamma) \langle s_i s_j \rangle + \frac{(\gamma_+ - \gamma_-)}{\mu N} \sum_m k_m \langle s_m s_i s_j \rangle \\ &\quad + 2\delta_{ij} \left\{ \beta + \gamma + \frac{(\gamma_+ - \gamma_-)}{\mu N} \sum_m k_m \langle s_m \rangle - \left[ \alpha + \frac{1}{2}(\gamma_+ - \gamma_-) \right] \langle s_i \rangle \right. \\ &\quad \left. - \frac{\gamma}{\mu N} \sum_m k_m \langle s_m s_i \rangle \right\}. \end{aligned} \quad (\text{S89})$$

Then we again make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , and obtain

$$\begin{aligned} \frac{d\langle s_i s_j \rangle}{dt} &= \alpha (\langle s_i \rangle + \langle s_j \rangle) + \frac{\gamma}{\mu N} \sum_m k_m \left[ \langle s_m s_i \rangle + \langle s_m s_j \rangle \right] - 2(\beta + \gamma) \langle s_i s_j \rangle \\ &\quad + 2\delta_{ij} \left[ \beta + \gamma - \alpha \langle s_i \rangle - \frac{\gamma}{\mu N} \sum_m k_m \langle s_i s_m \rangle \right]. \end{aligned} \quad (\text{S90})$$

Using Eqs. (S68), (S80) and (S90) we can then write a differential equation for elements of the correlation matrix,

$$\begin{aligned} \frac{d\Upsilon_{ij}}{dt} &= \frac{\gamma}{\mu N} \sum_m k_m (\Upsilon_{mi} + \Upsilon_{mj}) - 2(\beta + \gamma) \Upsilon_{ij} \\ &\quad + 2\delta_{ij} \left[ \beta + \gamma - \alpha \langle s_i \rangle - \frac{\gamma}{\mu N} \sum_m k_m \Upsilon_{im} - \langle s_i \rangle \frac{\gamma}{\mu N} \sum_m k_m \langle s_m \rangle \right]. \end{aligned} \quad (\text{S91})$$

Using the facts that  $\tau = \frac{\beta}{\gamma}$  and  $\langle s_i \rangle_{\text{st}} = \langle m \rangle_{\text{st}} = \frac{\alpha}{\beta}$ , from Eqs. (S46) and (S71) respectively,  $(\Upsilon_{ij})_{\text{st}}$  is found from Eq. (S91) to be

$$(\Upsilon_{ij})_{\text{st}} = \frac{\frac{1}{\mu N} \sum_m k_m [(\Upsilon_{mi})_{\text{st}} + (\Upsilon_{mj})_{\text{st}}] + 2\delta_{ij} \left[ (1 + \tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right) - \frac{1}{\mu N} \sum_m k_m (\Upsilon_{mi})_{\text{st}} \right]}{2(1 + \tau)}. \quad (\text{S92})$$

This equation for  $(\Upsilon_{ij})_{\text{st}}$  involves terms such as  $\sum_m k_m (\Upsilon_{mi})_{\text{st}}$  and from Eq. (S82) we know we ultimately need to calculate terms such as  $\sum_{ij} k_i k_j (\Upsilon_{ij})_{\text{st}}$ . Motivated by this and following the notation in [S4], we introduce a new variable  $S_x$ ,

$$S_x = \sum_{ij} k_i^x k_j (\Upsilon_{ij})_{\text{st}}, \quad (\text{S93})$$

where  $x$  is an integer such that  $k_i^x$  is simply the degree of node  $i$  raised to the  $x^{\text{th}}$  power. Substituting Eq. (S92) into Eq. (S93) allows us to write  $S_x$  as

$$S_x = \frac{S_x + \frac{1}{\mu} \bar{k}^x S_1 + 2N \bar{k}^{x+1} (1 + \tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right) - \frac{2}{\mu N} S_{x+1}}{2(1 + \tau)}, \quad (\text{S94})$$



where the overbar stands for an average over the degree distribution of the network,

$$\overline{f(k)} = \frac{1}{N} \sum_{i=1}^N f(k_i). \quad (\text{S95})$$

We then obtain the following recurrence relation,

$$\begin{aligned} S_{x+1} &= \frac{-\mu N}{2}(1+2\tau)S_x + \frac{N}{2} \left[ \overline{k^x} S_1 + 2\mu N \overline{k^{x+1}}(1+\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right) \right] \\ &\equiv \mathcal{C}S_x + \mathcal{D}_x, \end{aligned} \quad (\text{S96})$$

where,

$$\begin{aligned} \mathcal{C} &\equiv \frac{-\mu N}{2}(1+2\tau), \\ \mathcal{D}_x &\equiv \frac{N}{2} \left[ \overline{k^x} S_1 + 2\mu N \overline{k^{x+1}}(1+\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right) \right]. \end{aligned} \quad (\text{S97})$$

Eq. (S96) has the solution

$$\begin{aligned} S_{x+1} &= \mathcal{C}^x S_1 + \sum_{m=1}^x \mathcal{C}^{x-m} \mathcal{D}_m \\ &= \left[ \frac{-\mu N(1+2\tau)}{2} \right]^x S_1 \\ &\quad + \sum_{m=1}^x \left[ \frac{-\mu N(1+2\tau)}{2} \right]^{x-m} \frac{N}{2} \left[ \overline{k^m} S_1 + 2\mu N \overline{k^{m+1}}(1+\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right) \right]. \end{aligned} \quad (\text{S98})$$

Ultimately we want to determine  $S_1$ , as this is the term that appears in Eq. (S82). With this in mind, we can isolate the  $S_1$  term in Eq. (S98) by dividing both sides by  $\mathcal{C}^x$ , and taking the limit  $x \rightarrow \infty$ . The LHS is

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{S_{x+1}}{\left[ \frac{-\mu N(1+2\tau)}{2} \right]^x} &= \lim_{x \rightarrow \infty} \frac{\sum_{ij} k_i^{x+1} k_j \Upsilon_{ij}}{\left[ \frac{-\mu N(1+2\tau)}{2} \right]^x} \\ &= \left[ \frac{-\mu N(1+2\tau)}{2} \right] \lim_{x \rightarrow \infty} \sum_{ij} \left[ \frac{-2k_i}{\mu N(1+2\tau)} \right]^{x+1} k_j \Upsilon_{ij}. \end{aligned} \quad (\text{S99})$$

The limit on the right hand side will be zero provided

$$\left| \frac{-2k_i}{\mu N(1+2\tau)} \right| < 1. \quad (\text{S100})$$

The only possible ranges of  $\tau$  are  $\tau \geq 0$  and  $\tau \leq -2$  [see Sec. S8 B 2]. In each of these we have  $\left| \frac{1}{1+2\tau} \right| \leq 1$ , and thus the inequality in Eq. (S100) is fulfilled if  $k_i < \mu N/2$ . This is the case for  $\mu \geq 2$ . All networks we consider here have a mean degree higher than two.

We note that since  $k_i \leq N$ , the inequality in Eq. (S100) is actually always fulfilled for any  $\tau \neq -\frac{1}{2}$  provided that  $\mu$  is sufficiently large. But due to the restrictions on  $\tau$  [Sec. S8 B 2] we do not focus on such cases.

Having proven the limit in Eq. (S100) vanishes, we can find  $S_1$  from Eq. (S98) to be

$$S_1 = \frac{-\mu N^2(1+\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right) \sum_{m=1}^{\infty} \left[ \frac{-\mu N(1+2\tau)}{2} \right]^{-m} \overline{k^{m+1}}}{1 + \frac{N}{2} \sum_{m=1}^{\infty} \left[ \frac{-\mu N(1+2\tau)}{2} \right]^{-m} \overline{k^m}}. \quad (\text{S101})$$

Now notice that the summations in the above equation can be evaluated as follows,

$$\sum_{m=1}^{\infty} \mathcal{C}^{-m} \overline{k^{m+z}} = k^z \sum_{m=1}^{\infty} \overline{\mathcal{C}^{-m} k^m} = \frac{\overline{\mathcal{C}^{-1} k^{z+1}}}{1 - \overline{\mathcal{C}^{-1} k}}, \quad (\text{S102})$$

where we have used the infinite geometric series formula under the condition that  $|\overline{\mathcal{C}^{-1} k}| < 1$ , which is exactly the same condition as in Eq. (S100). After some manipulation we have the final form for  $S_1$ ,

$$S_1 = \frac{\mu N^2 (1 + \tau)(1 + 2\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right) \overline{\left(\frac{k^2}{(1+2\tau)\mu N + 2k}\right)}}{\tau + \frac{1}{\mu} \overline{\left(\frac{k^2}{(1+2\tau)\mu N + 2k}\right)}}. \quad (\text{S103})$$

This is then substituted back into Eq. (S82) and we find a parabolic relationship between  $\langle \sigma \rangle_{\text{st}}$  and  $\langle m \rangle_{\text{st}}$ ,

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{AA}}(\tau) \left(1 - \langle m \rangle_{\text{st}}^2\right), \quad (\text{S104})$$

with

$$H_{\text{AA}}(\tau) = 1 - \frac{(1 + \tau)(1 + 2\tau) \overline{\left(\frac{k^2}{(1+2\tau)\mu N + 2k}\right)}}{\tau \mu + \overline{\left(\frac{k^2}{(1+2\tau)\mu N + 2k}\right)}}. \quad (\text{S105})$$

The subscript ‘AA’ stands for ‘annealed approximation’. See Sec. S8 for detailed discussion on  $\tau$  and  $H(\tau)$ .

We see that in the limit  $\tau \rightarrow 0$ ,  $H_{\text{AA}}(\tau) \rightarrow 0$ . This matches what we observe in simulation. In the limit  $\tau \rightarrow \infty$  Eq. (S105) reduces to

$$\lim_{\tau \rightarrow \infty} H_{\text{AA}}(\tau) = 1 - \frac{\overline{k^2}}{\mu^2 N}. \quad (\text{S106})$$

Simulations in this limit give  $H(\tau) \rightarrow 1$ . Thus Eq. (S106) captures this up to corrections of order  $1/N$ .

### E. Comparison against simulations

In Fig. S3 we demonstrate how the approximation in Eq. (S105) compares to simulations for different size degree-regular networks (left), Erdős–Rényi networks (middle), and Barabási–Albert networks (right). As discussed above the theory matches simulation in the low  $\tau$  limit, and approximately so in the large  $\tau$  limit up to a  $\frac{1}{N}$  correction. There are discrepancies with simulation data for intermediate  $\tau$ . This is because in such a regime the structural properties of the network are most relevant. The nature of the annealed approximation is to replace the network with a weighted complete network, thus a lot of information about the original network structure is lost.

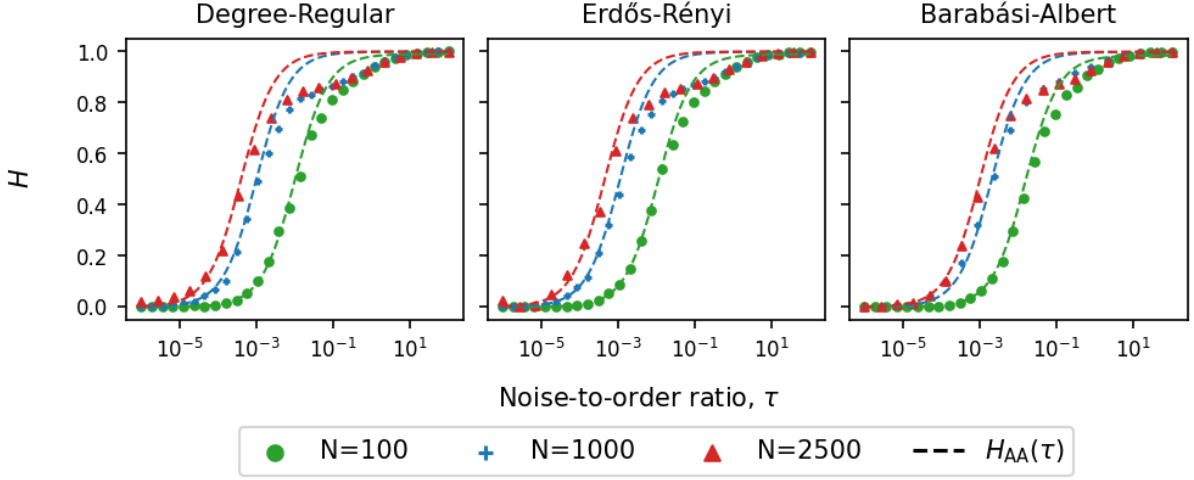


FIG. S3. Plots of  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)]$ , which characterises the scattering, against the noise-to-order ratio  $\tau$  for networks of size  $N$ . Degree-regular with  $\mu = 8$  (left), Erdős-Rényi with  $p = 8/N$  (middle), Barabási-Albert with 4 new links established upon adding a new node (right). For all three types of network we have  $\mu \approx 8$  (deviations are due to the finite size of the networks). The dashed lines are the results from the annealed approximation, Eq. (S105). Each marker is the result from averaging 1,000 independent simulations in the steady-state at a specific value of  $\tau$ . Model parameters for specific values of  $\tau$  are generated via the algorithm detailed in Sec. S10 B.

### S5. Analytical treatment of the model on an infinite square lattice

Next, we consider an infinite square lattice with  $N = L \times L$  sites. The calculation is based on that of [S1], but we correct a typo that previously prohibited solving the equations for a general choice of the model parameters. The notation in [S1] is different and a detailed discussion is given in Sec. S1 C.

#### A. Setup and spin flip rates

The state of the node at lattice site  $\mathbf{x} = (x_1, x_2)$  is written  $s(\mathbf{x}) \in \{-1, 1\}$ . The average state (over realisations of the dynamics) at site  $\mathbf{x}$  is written  $\langle s(\mathbf{x}) \rangle$ . The global mean magnetisation over the whole lattice is then  $\langle m \rangle = \sum_{\mathbf{x}} \langle s(\mathbf{x}) \rangle / N$ . The pair correlation between two spins  $s(\mathbf{x})$  and  $s(\mathbf{y})$  will be written  $\langle s(\mathbf{x})s(\mathbf{y}) \rangle$ . In summations we will use the notation  $\mathbf{y} \in \mathbf{x}^{(1)}$  to denote the set of von Neumann neighbours of  $\mathbf{x}$ , i.e. the set  $\{(x_1 - 1, x_2), (x_1 + 1, x_2), (x_1, x_2 - 1), (x_1, x_2 + 1)\}$ .

The spin flip probability (in discrete-time) is the probability with which the spin at site  $\mathbf{x}$  changes its state if it is selected for update. In continuous-time we can think of spin flip rates. The flip probability (or rate) for the spin at lattice site  $\mathbf{x}$  takes the form,

$$w(\mathbf{x}) = A(\mathbf{x}) + B(\mathbf{x}), \quad (\text{S107})$$

where  $A(\mathbf{x})$  is the contribution from the vertical process

$$\begin{aligned} A(\mathbf{x}) &= v'_+ \frac{1 - s(\mathbf{x})}{2} + v'_- \frac{1 + s(\mathbf{x})}{2} \\ &= \frac{1}{2} [v'_+ + v'_- - (v'_+ - v'_-)s(\mathbf{x})], \end{aligned} \quad (\text{S108})$$

and  $B(\mathbf{x})$  is the contribution from the horizontal process

$$\begin{aligned}
B(\mathbf{x}) &= \frac{1}{4} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \left[ h'_- \frac{1+s(\mathbf{x})}{2} \frac{1+s(\mathbf{y})}{2} + h_+ \frac{1-s(\mathbf{x})}{2} \frac{1+s(\mathbf{y})}{2} \right. \\
&\quad \left. + h'_+ \frac{1-s(\mathbf{x})}{2} \frac{1-s(\mathbf{y})}{2} + h_- \frac{1+s(\mathbf{x})}{2} \frac{1-s(\mathbf{y})}{2} \right] \\
&= \frac{1}{4} [(h_+ + h_-) + (h'_+ + h'_-)] - \frac{1}{4} [(h'_+ - h'_-) + (h_+ - h_-)] s(\mathbf{x}) \\
&\quad + \frac{1}{16} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} [(\gamma_+ - \gamma_-)s(\mathbf{y}) - (\gamma_+ + \gamma_-)s(\mathbf{x})s(\mathbf{y})]. \tag{S109}
\end{aligned}$$

The total spin flip probability is then

$$\begin{aligned}
w(\mathbf{x}) &= A(\mathbf{x}) + B(\mathbf{x}) \\
&= \frac{\eta}{2} - \frac{1}{2} \left\{ (v'_+ - v'_-) + \frac{1}{2} [(h'_+ - h'_-) + (h_+ - h_-)] \right\} s(\mathbf{x}) \\
&\quad + \frac{1}{16} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} [(\gamma_+ - \gamma_-)s(\mathbf{y}) - (\gamma_+ + \gamma_-)s(\mathbf{x})s(\mathbf{y})], \tag{S110}
\end{aligned}$$

where we have introduced expressions from Eq. (S4). We now make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , which allows us to re-write Eq. (S110) as

$$w(\mathbf{x}) = \frac{\eta}{2} - \frac{\alpha}{2} s(\mathbf{x}) - \frac{\gamma}{8} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} s(\mathbf{x})s(\mathbf{y}). \tag{S111}$$

## B. Magnetisation

We proceed with a discrete-time setup in mind. In a given time step, the spin at site  $\mathbf{x}$  changes state with probability  $\frac{w(\mathbf{x})}{N}$  where the  $\frac{1}{N}$  factor represents the probability of the spin being selected for update. If a spin flip occurs,  $s(\mathbf{x})$  changes by an amount  $-2s(\mathbf{x})$ . The change in the average spin,  $\langle s(\mathbf{x}) \rangle$ , is then

$$\langle s(\mathbf{x}, t + \Delta t) \rangle - \langle s(\mathbf{x}, t) \rangle = \frac{1}{N} \langle -2w(\mathbf{x})s(\mathbf{x}) \rangle, \tag{S112}$$

where  $\Delta t$  denotes a single time step. We now make the standard choice  $\Delta t = \frac{1}{N}$  and take the limit  $N \rightarrow \infty$  (i.e the continuous-time limit). Substituting in Eq. (S111) and evaluating the expectations leads to

$$\frac{d \langle s(\mathbf{x}) \rangle}{dt} = \alpha - \eta \langle s(\mathbf{x}) \rangle + \frac{\gamma}{4} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \langle s(\mathbf{y}) \rangle. \tag{S113}$$

The same result is obtained if we start out in continuous-time, and think of  $w(\mathbf{x})$  in Eq. (S111) as the rate for the spin at  $\mathbf{x}$  to flip.

Now we perform a summation over all lattice sites  $\mathbf{x}$  to find a differential equation for the magnetisation  $\langle m(t) \rangle$ ,

$$\frac{d \langle m(t) \rangle}{dt} = \alpha - \beta \langle m(t) \rangle. \tag{S114}$$

This is exactly the same as Eq. (S23) for the complete network, so the solution is identical.

## C. Interface density

To compute the interface density we start with the pair correlation  $\langle s(\mathbf{x})s(\mathbf{y}) \rangle$ . The quantity  $s(\mathbf{x})s(\mathbf{y})$  ( $\mathbf{x} \neq \mathbf{y}$ ) changes by an amount  $-2s(\mathbf{x})s(\mathbf{y})$  when either the spin at  $\mathbf{x}$  or  $\mathbf{y}$  flips. Working in the continuous-time limit,

we find

$$\begin{aligned} \frac{d\langle s(\mathbf{x})s(\mathbf{y}) \rangle}{dt} &= \langle -2[w(\mathbf{x}) + w(\mathbf{y})] s(\mathbf{x})s(\mathbf{y}) \rangle \\ &= -2\eta \langle s(\mathbf{x})s(\mathbf{y}) \rangle - 2\alpha \left[ \langle s(\mathbf{x}) \rangle + \langle s(\mathbf{y}) \rangle \right] - \frac{\gamma}{4} \left( \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} \langle s(\mathbf{y})s(\mathbf{z}) \rangle + \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \langle s(\mathbf{x})s(\mathbf{z}) \rangle \right). \end{aligned} \quad (\text{S115})$$

We now assume translational invariance so that  $\langle s(\mathbf{x})s(\mathbf{y}) \rangle$  is a function of  $\mathbf{x} - \mathbf{y}$  only. We write  $\langle s(\mathbf{x})s(\mathbf{y}) \rangle = C(\mathbf{x} - \mathbf{y})$ . For  $\mathbf{r} \equiv \mathbf{x} - \mathbf{y} \neq 0$ , we then have

$$\frac{dC(\mathbf{r})}{dt} = 2\beta \left[ \langle m \rangle_{\text{st}} (\langle s(\mathbf{x}) \rangle + \langle s(\mathbf{y}) \rangle) - C(\mathbf{r}) \right] + 2\gamma \Delta C(\mathbf{r}), \quad (\text{S116})$$

where  $\Delta$  is the lattice Laplacian,

$$\Delta f(\mathbf{x}) = -f(\mathbf{x}) + \frac{1}{4} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} f(\mathbf{y}). \quad (\text{S117})$$

All lattice sites are statistically equivalent to one another in the ensemble of realisations (assuming initial conditions are independent and identical for each lattice site). In the stationary state we then have  $\langle s(\mathbf{x}) \rangle_{\text{st}} = \langle s(\mathbf{y}) \rangle_{\text{st}} = \langle m \rangle_{\text{st}}$ . We also have  $\frac{dC(\mathbf{r})}{dt} = 0$ , so Eq. (S116) becomes

$$0 = \beta \left[ \langle m \rangle_{\text{st}}^2 - C(\mathbf{r}) \right] + \gamma \Delta C(\mathbf{r}), \quad (\text{S118})$$

which has the same form as equation (S75) in Ref. [S1], but where  $2\lambda$  in the expression in the reference is now replaced by  $p'_I + p'_E$  if we were to map back to the model parameters of that earlier reference via Eq. (S8). We note a typographical error in the calculation in [S1]: In equation (S59) of that reference the final term in the last line is  $s(\mathbf{x}')$ , whereas it should be  $s(\mathbf{x})$ . Correcting this means that a problematic term which could not be dealt with in this earlier work no longer appears in Eqs. (S113) and (S115) of the present paper. In Ref. [S1], the problematic term is removed by making the assumption  $p'_I = p'_E$ , but the present calculation shows that this assumption is not actually required to proceed.

We omit the remainder of the derivation as it follows that in Ref. [S1]. We find that the expression for the average stationary-state interface is given by

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{SL}}(\tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right), \quad (\text{S119})$$

the subscript ‘SL’ stands for ‘square lattice’. This was obtained for a restricted set of model parameters in [S1] but we now know this to be true when all model parameters are in free variation. The function  $H_{\text{SL}}(\tau)$  is given by

$$H_{\text{SL}}(\tau) = \frac{\pi(1+\tau)}{2K\left(\frac{1}{1+\tau}\right)} - \tau. \quad (\text{S120})$$

$K(\cdot)$  is the complete elliptical integral of the first kind defined as

$$K(x) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1 - x^2 \sin^2(\theta)}}. \quad (\text{S121})$$

The noise-to-order ratio  $\tau$  is defined as in Eq. (S46),

$$\tau = \frac{\beta}{\gamma} = \frac{(h'_+ + h'_-) + (v'_+ + v'_-)}{\frac{1}{2} [(h_+ - h'_+) + (h_- - h'_-)]}. \quad (\text{S122})$$

See Sec. S8 for detailed discussion on  $\tau$  and  $H(\tau)$ .

We note that  $K(x)$  is symmetric about  $x = 0$  and defined only for  $|x| \leq 1$ . When  $x \rightarrow 0$ ,  $K(x) \rightarrow \frac{\pi}{2}$ , and when  $x \rightarrow \pm 1$ ,  $K(x) \rightarrow \infty$ . The argument in Eq. (S120) is  $\frac{1}{1+\tau}$ , thus  $K$  is only defined for  $\tau \leq -2$  and  $\tau \geq 0$ . A further discussion on the different limits of  $\tau$  can be found in Sec. S8 C.

Fig. S4 confirms the validity of these results in simulations, and demonstrates that Eq. (S119) holds without the restrictions on model parameters that were needed in [S1].

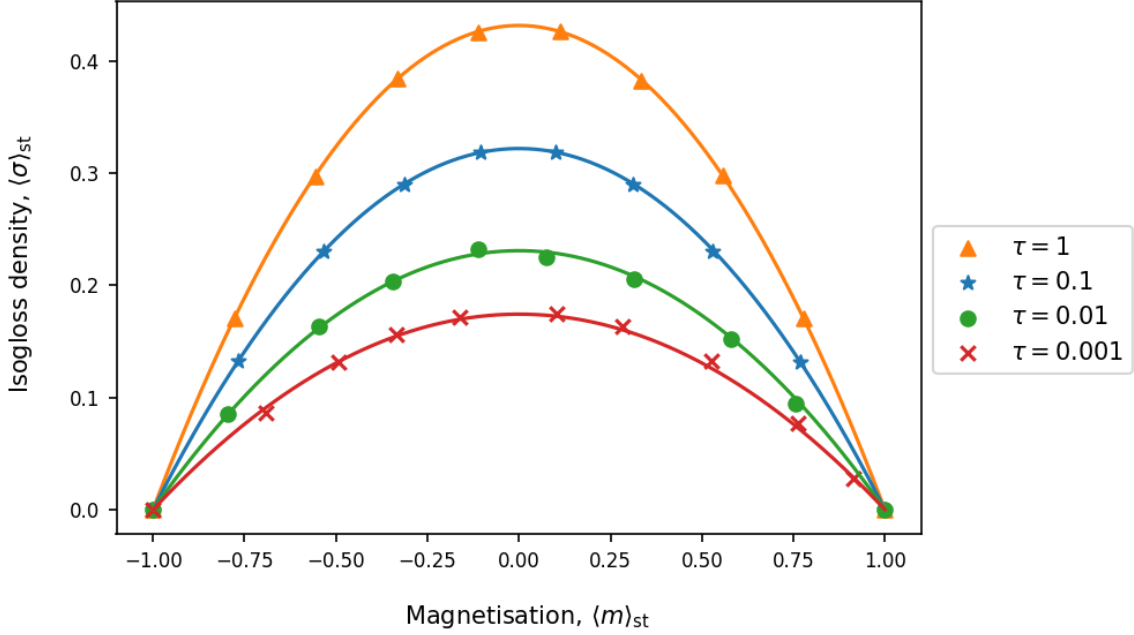


FIG. S4. Plots of  $\langle \sigma \rangle_{st}$  against  $\langle m \rangle_{st}$ . Solid lines are the analytic solutions,  $H_{SL}(\tau)$ , from Eq. (S119) for different values of the noise-to-order ratio,  $\tau$ . These solutions apply for infinite 2D square lattices. Each marker is the result from averaging 100 independent simulations which were performed on 2D square lattices with  $N = 10,000$  and periodic boundary conditions. Model parameters for a given  $\langle m \rangle_{st}$  and  $\tau$  can be generated via the method in Sec. S10B. The behaviour of the model in various limits of  $\tau$  is discussed in Secs. S8B and S8C. In particular the height of the parabola tends to  $\frac{1}{2}$  as  $\tau \rightarrow \infty$ .

## S6. Approach based on network walks

### A. Useful definitions and identities for the further analysis

#### 1. Definitions

For any non-negative integer  $\ell$  we define a walk of length  $\ell$  as an ordered set of  $\ell + 1$  nodes  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{\ell+1}$  on the network, such that  $\mathbf{x}_{i+1}$  is a nearest neighbour of  $\mathbf{x}_i$  for all  $i = 1, \dots, \ell$ . We do not include networks with self-loops or multi-links, as these are not relevant in the context of the model. We note that the  $\ell + 1$  nodes do not need to be pairwise different. That is to say, the walker can visit the same node on the network multiple times.

An illustration of a simple network with five nodes is shown in Fig. S5. There are four walks of length  $\ell = 2$  starting at node 1. These are  $(1, 2, 4)$ ,  $(1, 3, 4)$ ,  $(1, 2, 1)$  and  $(1, 3, 1)$ . Similarly, there are two walks of length  $\ell = 2$  starting at node 1 and ending at node 4, namely  $(1, 2, 4)$  and  $(1, 3, 4)$ . As a further example, there are four walks of length  $\ell = 3$  starting at 1 and ending at 2, these are  $(1, 2, 1, 2)$ ,  $(1, 2, 4, 2)$ ,  $(1, 3, 4, 2)$  and  $(1, 3, 1, 2)$ .

We define

$$W^{(\ell)}(\mathbf{x}, \mathbf{y}) = \text{number of distinct walks of length } \ell \text{ starting at } \mathbf{x} \text{ and ending at } \mathbf{y},$$

$$\nu^{(\ell)}(\mathbf{x}) = \text{number of distinct walks of length } \ell \text{ starting at } \mathbf{x}. \quad (\text{S123})$$

For any node  $\mathbf{x}$  we also introduce

$$\mathbf{x}^{(\ell)} = \{\mathbf{y} \mid \mathbf{y} \text{ is the endpoint of a walk of length } \ell \text{ starting at } \mathbf{x}\}. \quad (\text{S124})$$



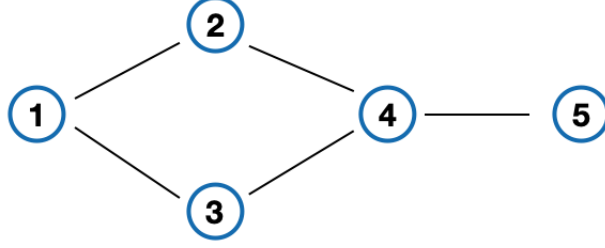


FIG. S5. Illustration of walks on a simple network (see text).

We note that  $\mathbf{x}^{(\ell)}$  is to be understood as a set, i.e. no node can appear in  $\mathbf{x}^{(\ell)}$  multiple times. Using again the example in Fig. S5, and setting  $\mathbf{x}$  to be node 1, we have

$$\begin{aligned}\mathbf{x}^{(1)} &= \{2, 3\}, \\ \mathbf{x}^{(2)} &= \{1, 4\}, \\ \mathbf{x}^{(3)} &= \{2, 3, 5\}, \\ \mathbf{x}^{(4)} &= \{1, 4\}.\end{aligned}\tag{S125}$$

We then have

$$\nu^{(\ell)}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}).\tag{S126}$$

We also note that  $\nu^{(1)}(\mathbf{x})$  is simply the number of nearest neighbours of node  $\mathbf{x}$ , i.e. its degree.

The quantity  $W^{(\ell)}(\mathbf{x}, \mathbf{y})$  is given by the  $\mathbf{xy}$  element of the  $\ell^{\text{th}}$  power of the adjacency matrix  $\mathbf{A}$  of the network,

$$W^{(\ell)}(\mathbf{x}, \mathbf{y}) = (A^\ell)_{\mathbf{xy}}.\tag{S127}$$

This can be seen by realising that

$$(A^\ell)_{\mathbf{xy}} = \sum_{\mathbf{z}_1, \dots, \mathbf{z}_{\ell-1}} A_{\mathbf{xz}_1} A_{\mathbf{z}_1 \mathbf{z}_2} \dots A_{\mathbf{z}_{\ell-1} \mathbf{y}},\tag{S128}$$

and keeping in mind that elements of the adjacency matrix only take values zero or one. Consequently we also have

$$\nu^{(\ell)}(\mathbf{x}) = \sum_{\mathbf{y}} (A^\ell)_{\mathbf{xy}}.\tag{S129}$$

## 2. Identities

There are several identities we will use throughout in order to manipulate the summations. Recall, as per Sec. S1 A, we only consider undirected networks.

The first identity is

$$\begin{aligned}
\sum_{\mathbf{x}} f(\mathbf{x}) \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} g(\mathbf{y}) &= \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} f(\mathbf{x})g(\mathbf{y}) \\
&= \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \ell\text{-connected}}} f(\mathbf{x})g(\mathbf{y}) \\
&= \sum_{\mathbf{y}} \sum_{\mathbf{x} \in \mathbf{y}^{(\ell)}} f(\mathbf{x})g(\mathbf{y}) \\
&= \sum_{\mathbf{x}} g(\mathbf{x}) \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} f(\mathbf{y}), \tag{S130}
\end{aligned}$$

where we say that two nodes  $\mathbf{x}$  and  $\mathbf{y}$  are ‘ $\ell$ -connected’ if there exists a walk of length  $\ell$  connecting  $\mathbf{x}$  and  $\mathbf{y}$ , i.e. if  $\mathbf{y} \in \mathbf{x}^{(\ell)}$  or equivalently,  $\mathbf{x} \in \mathbf{y}^{(\ell)}$ . In the last line we swap the labels  $\mathbf{x}$  and  $\mathbf{y}$ . This identity says that if we have a sum of a function over all nodes, and then an inner sum of a different function over the  $\ell$  nearest neighbours of those nodes, we can swap the functions around.

A similar identity, derived in the same way, valid for any function  $f(\mathbf{x}, \mathbf{y})$ ,

$$\sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} f(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} f(\mathbf{y}, \mathbf{x}). \tag{S131}$$

A second identity is derived by first noting that for any fixed  $\mathbf{x}$  the following two statements for two nodes  $\mathbf{y}$  and  $\mathbf{z}$  are equivalent:

- (1)  $\mathbf{y} \in \mathbf{x}^{(\ell)}$  and  $\mathbf{z} \in \mathbf{y}^{(1)}$ ;
- (2)  $\mathbf{z} \in \mathbf{x}^{(\ell+1)}$  and  $\mathbf{y} \in \mathbf{x}^{(\ell)} \cap \mathbf{z}^{(1)}$ .

Statement (2) implies that  $\mathbf{y} \in \mathbf{x}^{(\ell)}$  and that  $\mathbf{z}$  and  $\mathbf{y}$  are nearest neighbours, so statement (1) follows. Further, if  $\mathbf{y} \in \mathbf{x}^{(\ell)}$ , i.e. there is a walk of length  $\ell$  from  $\mathbf{x}$  to  $\mathbf{y}$ , and if  $\mathbf{z}$  is a nearest neighbour of  $\mathbf{y}$ , then there exists a walk of length  $\ell + 1$  from  $\mathbf{x}$  to  $\mathbf{z}$ . This means that (2) is fulfilled. Using this equivalence we have

$$\begin{aligned}
\sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} f(\mathbf{x}, \mathbf{y})g(\mathbf{x}, \mathbf{z}) &= \sum_{\mathbf{x}} \sum_{\mathbf{z} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)} \cap \mathbf{z}^{(1)}} f(\mathbf{x}, \mathbf{y}) \right) g(\mathbf{x}, \mathbf{z}) \\
&= \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{x}^{(\ell)} \cap \mathbf{y}^{(1)}} f(\mathbf{x}, \mathbf{z}) \right) g(\mathbf{x}, \mathbf{y}), \tag{S132}
\end{aligned}$$

for all functions  $f(\cdot, \cdot)$  and  $g(\cdot, \cdot)$ .

In the same way the following two statements are equivalent to one another,

- (1)  $\mathbf{y} \in \mathbf{x}^{(\ell)}$  and  $\mathbf{z} \in \mathbf{x}^{(1)}$ ;
- (2)  $\mathbf{y} \in \mathbf{z}^{(\ell+1)}$  and  $\mathbf{x} \in \mathbf{y}^{(\ell)} \cap \mathbf{z}^{(1)}$ ,

leading to the identity,

$$\begin{aligned}
\sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} f(\mathbf{x}, \mathbf{y})g(\mathbf{y}, \mathbf{z}) &= \sum_{\mathbf{z}} \sum_{\mathbf{y} \in \mathbf{z}^{(\ell+1)}} \left( \sum_{\mathbf{x} \in \mathbf{y}^{(\ell)} \cap \mathbf{z}^{(1)}} f(\mathbf{x}, \mathbf{y}) \right) g(\mathbf{y}, \mathbf{z}) \\
&= \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{y}^{(\ell)} \cap \mathbf{x}^{(1)}} f(\mathbf{z}, \mathbf{y}) \right) g(\mathbf{y}, \mathbf{x}), \tag{S133}
\end{aligned}$$

for all functions  $f(\cdot, \cdot)$  and  $g(\cdot, \cdot)$ .

### B. Magnetisation

The spin-flip probability,  $w(\mathbf{x})$ , in the case of a general network is constructed in exactly the same way as that of the infinite square lattice in Sec. S5. The only difference is that the pre-factor 1/4 in Eq. (S109) becomes  $1/\nu^{(1)}(\mathbf{x})$ . This leads to an equation for  $w(\mathbf{x})$  analogous to Eq. (S111),

$$w(\mathbf{x}) = \frac{\eta}{2} - \frac{\alpha}{2}s(\mathbf{x}) - \frac{\gamma}{2} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} s(\mathbf{x})s(\mathbf{y}), \quad (\text{S134})$$

where we have used the shorthands from Eq. (S4). We note that we have already made the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ . Using this we can write an equation which is analogous to Eq. (S113),

$$\frac{dS(\mathbf{x})}{dt} = \alpha - \eta S(\mathbf{x}) + \gamma \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}). \quad (\text{S135})$$

We then take a sum over all nodes  $\mathbf{x}$  and find the following differential equation for the magnetisation,

$$\begin{aligned} \frac{d\langle m \rangle}{dt} &= \alpha - \eta \langle m \rangle + \gamma \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}) \\ &= \alpha - \eta \langle m \rangle + \gamma \frac{1}{N} \sum_{\mathbf{x}} S(\mathbf{x}) \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})}, \end{aligned} \quad (\text{S136})$$

where we have made use of Eq. (S130).

For each node  $\mathbf{x}$  we introduce

$$\lambda_{\mathbf{x},1} = \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})}, \quad (\text{S137})$$

and we further define

$$\langle m_1 \rangle = \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} S(\mathbf{x}). \quad (\text{S138})$$

We can write Eq. (S136) as

$$\begin{aligned} \frac{d\langle m \rangle}{dt} &= \alpha - \eta \langle m \rangle + \gamma \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} S(\mathbf{x}) \\ &= \alpha - \eta \langle m \rangle + \gamma \langle m_1 \rangle. \end{aligned} \quad (\text{S139})$$

Next, using Eq. (S135), we have

$$\frac{d\langle m_1 \rangle}{dt} = \alpha \Lambda_1 - \eta \langle m_1 \rangle + \gamma \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}), \quad (\text{S140})$$

where we have defined

$$\Lambda_1 = \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1}. \quad (\text{S141})$$

We can make the following simplification,

$$\begin{aligned}
\Lambda_1 &= \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})} \\
&= \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} 1 \\
&= \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(1)}(\mathbf{x})} \cdot \nu^{(1)}(\mathbf{x}) \\
&= 1,
\end{aligned} \tag{S142}$$

where we have used Eq. (S130) in going from the second to third line. Thus, we have

$$\begin{aligned}
\frac{d\langle m_1 \rangle}{dt} &= \alpha - \eta \langle m_1 \rangle + \gamma \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}) \\
&= \alpha - \eta \langle m_1 \rangle + \gamma \langle m_2 \rangle,
\end{aligned} \tag{S143}$$

with the definition

$$\langle m_2 \rangle \equiv \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}). \tag{S144}$$

This expression can be simplified as follows,

$$\begin{aligned}
\langle m_2 \rangle &= \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},1} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} S(\mathbf{y}) \\
&= \frac{1}{N} \sum_{\mathbf{x}} S(\mathbf{x}) \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})} \lambda_{\mathbf{y},1} \\
&= \frac{1}{N} \sum_{\mathbf{x}} S(\mathbf{x}) \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{z})} \\
&\equiv \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},2} S(\mathbf{x}),
\end{aligned} \tag{S145}$$

where we again have used Eq. (S130) in going from the first to second line, and where we have defined

$$\lambda_{\mathbf{x},2} \equiv \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{z})}. \tag{S146}$$

We now make the following recursive definition:

$$\lambda_{\mathbf{x},n} = \begin{cases} \sum_{\mathbf{x}_1 \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{x}_1)} \lambda_{\mathbf{x}_1, n-1} & n = 1, 2, 3, \dots \\ 1 & n = 0 \end{cases}. \tag{S147}$$

Further we write

$$\langle m_n \rangle \equiv \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},n} S(\mathbf{x}). \tag{S148}$$

We also define

$$\Lambda_n \equiv \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},n}, \tag{S149}$$

and note that

$$\begin{aligned}
\Lambda_n &= \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},n} \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{x}_1 \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{x}_1)} \lambda_{\mathbf{x}_1, n-1} \\
&= \frac{1}{N} \sum_{\mathbf{x}} \frac{\lambda_{\mathbf{x}, n-1}}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{x}_1 \in \mathbf{x}^{(1)}} 1 \\
&= \frac{1}{N} \sum_{\mathbf{x}} \frac{\lambda_{\mathbf{x}, n-1}}{\nu^{(1)}(\mathbf{x})} \cdot \nu^{(1)}(\mathbf{x}) \\
&= \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x}, n-1} = \Lambda_{n-1}.
\end{aligned} \tag{S150}$$

Thus, all  $\Lambda_n$  take the same value, and using Eq. (S142), we find  $\Lambda_n = 1$  for all  $n$ .

From Eq. (S135) we can derive differential equations for general  $\langle m_n \rangle$ ,  $n = 1, 2, 3, \dots$ , leading to the following hierarchy,

$$\frac{d \langle m_n \rangle}{dt} = \alpha - \eta \langle m_n \rangle + \gamma \langle m_{n+1} \rangle. \tag{S151}$$

Next, we take the steady-state limit, and multiply the equation through by  $\frac{\gamma^n}{\eta^n}$ , and sum both sides over  $n$  (from 0 to  $\infty$ ),

$$\begin{aligned}
0 &= \alpha \sum_{n=0}^{\infty} \frac{\gamma^n}{\eta^n} - \sum_{n=0}^{\infty} \left( \frac{\gamma^n}{\eta^{n-1}} \langle m_n \rangle_{\text{st}} - \frac{\gamma^{n+1}}{\eta^n} \langle m_{n+1} \rangle_{\text{st}} \right) \\
&= \alpha \frac{1}{1 - \frac{\gamma}{\eta}} - \eta \langle m \rangle_{\text{st}} + \lim_{n \rightarrow \infty} \left( \frac{\gamma^{n+1}}{\eta^n} \langle m_{n+1} \rangle_{\text{st}} \right).
\end{aligned} \tag{S152}$$

The first term results from an infinite geometric series, which only converges provided  $|\gamma/\eta| = |1/(1 + \tau)| < 1$  [where we have used the definitions in Eqs. (S4) and (S46)]. This is only the case provided that  $\tau < -2$  or  $\tau > 0$ . The second and third terms in the second line of Eq. (S152) result from the telescopic sum in the first line. We note that  $\langle m_0 \rangle = \langle m \rangle$  by construction.

The limit in the second line of Eq. (S152) vanishes. Recall that all  $\lambda_{\mathbf{x},n} \geq 0$  and  $\Lambda_n \equiv \frac{1}{N} \sum_{\mathbf{x}} \lambda_{\mathbf{x},n} = 1$ , so we have  $0 \leq \lambda_{\mathbf{x},n} \leq 1$ . Then from the definition in Eq. (S148), we always have  $-1 \leq \langle m_n \rangle_{\text{st}} \leq 1$ . Since  $|\gamma/\eta| < 1$  the limit then vanishes.

Ultimately then we can solve for  $\langle m \rangle_{\text{st}}$ . We find

$$\langle m \rangle_{\text{st}} = \frac{\alpha}{\eta - \gamma} = \frac{\alpha}{\beta}, \tag{S153}$$

where  $\alpha$  and  $\beta$  are given in Eq. (S4).

This coincides with the result we obtained for infinite complete networks, Eq. (S24), using the pair approximation [see Sec. S3B], and using the annealed approximation, Eq. (S71). We note that the derivation in the current section did not require any approximation, and is therefore valid for any finite undirected network.

### C. Generalisation to ‘weighted magnetisation’

It is possible to generalise the procedure in the previous subsection. Consider the following ‘weighted’ magnetisation,

$$\langle m_z \rangle = \frac{\frac{1}{N} \sum_{\mathbf{x}} z_{\mathbf{x}} S(\mathbf{x})}{\frac{1}{N} \sum_{\mathbf{x}} z_{\mathbf{x}}}, \tag{S154}$$

where the  $z_{\mathbf{x}} \geq 0$  are site dependent weightings. We define

$$\langle m_{z,n} \rangle = \frac{\frac{1}{N} \sum_{\mathbf{x}} z_{\mathbf{x},n} S(\mathbf{x})}{\frac{1}{N} \sum_{\mathbf{x}} z_{\mathbf{x},n}}, \quad (\text{S155})$$

with

$$z_{\mathbf{x},n} = \begin{cases} \sum_{\mathbf{x}_1 \in \mathbf{x}^{(1)}} \frac{1}{\nu^{(1)}(\mathbf{x}_1)} z_{\mathbf{x}_1, n-1} & n = 1, 2, 3, \dots \\ z_{\mathbf{x}} & n = 0 \end{cases}. \quad (\text{S156})$$

Similar to Eq. (S142) we then have

$$\zeta_n \equiv \frac{1}{N} \sum_{\mathbf{x}} z_{\mathbf{x},n} = 1 \quad \forall n. \quad (\text{S157})$$

Thus the differential equation for  $\langle m_{z,n} \rangle$  follows the same recurrence relation as Eq. (S151),

$$\frac{d \langle m_{z,n} \rangle}{dt} = \alpha - \eta \langle m_{z,n} \rangle + \gamma \langle m_{z,n+1} \rangle. \quad (\text{S158})$$

We can solve this in the same way as before, leading to

$$\langle m_z \rangle_{\text{st}} = \langle m \rangle_{\text{st}}. \quad (\text{S159})$$

Thus the steady-state ‘weighted’ magnetisation is always the same, regardless of the weighting. This means, for example, that the magnetisation of a group of specific degree nodes will have the same steady-state independent of the degree chosen. This will become relevant in Sec. S7.

#### D. Interface density

It turns out to be convenient to introduce the following object,

$$S^{(\ell)}(\mathbf{x}) = \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \langle s(\mathbf{x}) s(\mathbf{y}) \rangle. \quad (\text{S160})$$

Broadly speaking, this is the correlation function of  $s(\mathbf{x})$  with spins at sites  $\mathbf{y}$  that can be reached from  $\mathbf{x}$  in walks of length  $\ell$ . Each  $\mathbf{y} \in \mathbf{x}^{(\ell)}$  is weighted by the number of walks of length  $\ell$  connecting  $\mathbf{x}$  and  $\mathbf{y}$ . Noting the definition of  $\nu^{(\ell)}(\mathbf{x})$  from Eq. (S126), the pre-factor  $1/\nu^{(\ell)}(\mathbf{x})$  ensures overall normalisation, such that  $-1 \leq S^{(\ell)}(\mathbf{x}) \leq 1$ .

As in Sec. S5C, we use again the fact that  $s(\mathbf{x})s(\mathbf{y})$  changes by an amount  $-2s(\mathbf{x})s(\mathbf{y})$  when either  $s(\mathbf{x})$  or  $s(\mathbf{y})$  flip ( $\mathbf{x} \neq \mathbf{y}$ ). Noting that the node  $\mathbf{x}$  itself may be contained in the set  $\mathbf{x}^{(\ell)}$  for some  $\ell$ , and recognising that  $s(\mathbf{x})^2$  always takes the value one (and hence does not change if the spin at  $\mathbf{x}$  flips), we find in the continuous-time limit

$$\frac{dS^{(\ell)}(x)}{dt} = \frac{4W^{(\ell)}(\mathbf{x}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{x})} \langle w(\mathbf{x}) \rangle - \frac{2}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \langle s(\mathbf{x}) s(\mathbf{y}) [w(\mathbf{x}) + w(\mathbf{y})] \rangle. \quad (\text{S161})$$

The first term corrects for any contribution from  $\mathbf{y} = \mathbf{x}$  in the second term (in the event that  $\mathbf{x} \in \mathbf{x}^{(\ell)}$ ).

We will now use Eq. (S134), which we repeat here for convenience,

$$w(\mathbf{x}) = \frac{\eta}{2} - \frac{\alpha}{2} s(\mathbf{x}) - \frac{\gamma}{2} \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(1)}} s(\mathbf{x}) s(\mathbf{y}), \quad (\text{S162})$$

and the definition in Eq. (S160). We find

$$\langle w(\mathbf{x}) \rangle = \frac{1}{2} \left[ \eta - \alpha S(\mathbf{x}) - \gamma S^{(1)}(\mathbf{x}) \right]. \quad (\text{S163})$$

We also have

$$\begin{aligned}
& \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \langle s(\mathbf{x})s(\mathbf{y}) [w(\mathbf{x}) + w(\mathbf{y})] \rangle \\
&= \eta \nu^{(\ell)}(\mathbf{x}) S^{(\ell)}(\mathbf{x}) - \frac{\alpha}{2} \left[ \nu^{(\ell)}(\mathbf{x}) S(\mathbf{x}) + \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) S(\mathbf{y}) \right] \\
&\quad - \frac{\gamma}{2} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \left[ \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \langle s(\mathbf{x})s(\mathbf{z}) \rangle + \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} \langle s(\mathbf{z})s(\mathbf{y}) \rangle \right]. \tag{S164}
\end{aligned}$$

Eq. (S161) then becomes

$$\begin{aligned}
\frac{dS^{(\ell)}(\mathbf{x})}{dt} &= \frac{2W^{(\ell)}(\mathbf{x}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{x})} \left[ \eta - \alpha S(\mathbf{x}) - \gamma S^{(1)}(\mathbf{x}) \right] \\
&\quad - 2\eta S^{(\ell)}(\mathbf{x}) + \alpha \left[ S(\mathbf{x}) + \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) S(\mathbf{y}) \right] \\
&\quad + \frac{\gamma}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \left[ \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \langle s(\mathbf{x})s(\mathbf{z}) \rangle + \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} \langle s(\mathbf{z})s(\mathbf{y}) \rangle \right]. \tag{S165}
\end{aligned}$$

In the steady-state, the derivative on the left-hand side vanishes. We can then divide through by  $\gamma$  so that the model parameters only appear in the following combinations,

$$\frac{\eta}{\gamma} = 1 + \tau, \tag{S166a}$$

$$\frac{\alpha}{\gamma} = \tau \langle m \rangle_{\text{st}}, \tag{S166b}$$

where  $\langle m \rangle_{\text{st}}$  and  $\tau$  are as in Eqs. (S24) and (S46) respectively. As we will show below, this means that the steady-state interface density can also be determined solely from  $\langle m \rangle_{\text{st}}$  and  $\tau$ .

We now sum both sides of Eq. (S165) over all nodes  $\mathbf{x}$  in the network to derive an equation for the overall correlation  $C^{(\ell)}$ , defined as

$$C^{(\ell)} \equiv \frac{1}{N} \sum_{\mathbf{x}} S^{(\ell)}(\mathbf{x}). \tag{S167}$$

We find,

$$\begin{aligned}
\frac{dC^{(\ell)}}{dt} &= 2 \frac{1}{N} \sum_{\mathbf{x}} \left\{ \frac{W^{(\ell)}(\mathbf{x}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{x})} \left[ \eta - \alpha S(\mathbf{x}) - \gamma S^{(1)}(\mathbf{x}) \right] \right\} - 2\eta C^{(\ell)} \\
&\quad + \alpha \left[ \langle m \rangle + \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) S(\mathbf{y}) \right] \\
&\quad + \frac{\gamma}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \left[ \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \langle s(\mathbf{x})s(\mathbf{z}) \rangle \right. \\
&\quad \quad \quad \left. + \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} \langle s(\mathbf{z})s(\mathbf{y}) \rangle \right]. \tag{S168}
\end{aligned}$$

To proceed we introduce a number of shorthands. We write

$$\Omega_{\mathbf{x}}^{(\ell)} \equiv \frac{W^{(\ell)}(\mathbf{x}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{x})}, \quad (\text{S169a})$$

$$\chi_{\mathbf{x}}^{(\ell)} \equiv \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \frac{W^{(\ell)}(\mathbf{y}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{y})}, \quad (\text{S169b})$$

$$\psi_{\mathbf{xy}}^{(\ell+1)} \equiv \sum_{\mathbf{z} \in \mathbf{x}^{(\ell)} \cap \mathbf{y}^{(1)}} \left[ \frac{W^{(\ell)}(\mathbf{x}, \mathbf{z}) W^{(1)}(\mathbf{z}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{x}) \nu^{(1)}(\mathbf{z})} \right], \quad (\text{S169c})$$

$$\phi_{\mathbf{xy}}^{(\ell+1)} \equiv \sum_{\mathbf{z} \in \mathbf{y}^{(\ell)} \cap \mathbf{x}^{(1)}} \left[ \frac{W^{(\ell)}(\mathbf{z}, \mathbf{y}) W^{(1)}(\mathbf{z}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{z}) \nu^{(1)}(\mathbf{z})} \right]. \quad (\text{S169d})$$

In practice these coefficients can be evaluated from the  $N \times N$  adjacency matrix  $\mathbf{A}$  of the network. We here recall that the  $\mathbf{xy}$  element of the  $\ell^{\text{th}}$  power of the adjacency gives the number of possible length  $\ell$  walks starting at  $\mathbf{x}$  and ending at  $\mathbf{y}$ , denoted  $W^{(\ell)}(\mathbf{x}, \mathbf{y})$ , see Eq. (S127). The sum of such elements over all  $\mathbf{y}$  gives the total number of walks of length  $\ell$  starting from  $\mathbf{x}$ , denoted  $\nu^{(\ell)}(\mathbf{x})$ , see Eq. (S129). Using these, we have

$$\Omega_{\mathbf{x}}^{(\ell)} = \frac{(A^\ell)_{\mathbf{xx}}}{\sum_{\mathbf{y}} (A^\ell)_{\mathbf{xy}}}, \quad (\text{S170a})$$

$$\chi_{\mathbf{x}}^{(\ell)} = \sum_{\mathbf{y}} \frac{(A^\ell)_{\mathbf{yx}}}{\sum_{\mathbf{z}} (A^\ell)_{\mathbf{yz}}}, \quad (\text{S170b})$$

$$\psi_{\mathbf{xy}}^{(\ell+1)} = \frac{1}{\sum_{\mathbf{z}} (A^\ell)_{\mathbf{xz}}} \left[ \sum_{\mathbf{z}} (A^\ell)_{\mathbf{xz}} \frac{1}{\sum_{\mathbf{u}} A_{\mathbf{zu}}} A_{\mathbf{zy}} \right], \quad (\text{S170c})$$

$$\phi_{\mathbf{xy}}^{(\ell+1)} = \sum_{\mathbf{z}} \left[ \frac{(A^\ell)_{\mathbf{zy}} A_{\mathbf{zx}}}{\sum_{\mathbf{u}} (A^k)_{\mathbf{zu}} \sum_{\mathbf{u}} A_{\mathbf{zu}}} \right]. \quad (\text{S170d})$$

We now proceed to write Eq. (S168) in more compact form. Using Eq. (S131) we observe for the term on the second line,

$$\begin{aligned} \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) S(\mathbf{y}) &= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \left( \frac{W^{(\ell)}(\mathbf{x}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{x})} S(\mathbf{y}) \right) \\ &= \frac{1}{N} \sum_{\mathbf{x}} \left( \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \frac{W^{(\ell)}(\mathbf{y}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{y})} \right) S(\mathbf{x}) \\ &= \frac{1}{N} \sum_{\mathbf{x}} \chi_{\mathbf{x}}^{(\ell)} S(\mathbf{x}). \end{aligned} \quad (\text{S171})$$

For the first term on the third line,

$$\begin{aligned} \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \frac{1}{\nu^{(1)}(\mathbf{y})} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \langle s(\mathbf{x}) s(\mathbf{z}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \frac{W^{(\ell)}(\mathbf{x}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{x}) \nu^{(1)}(\mathbf{y})} \langle s(\mathbf{x}) s(\mathbf{z}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{x}^{(\ell)} \cap \mathbf{y}^{(1)}} \frac{W^{(\ell)}(\mathbf{x}, \mathbf{z})}{\nu^{(\ell)}(\mathbf{x}) \nu^{(1)}(\mathbf{z})} \right) \langle s(\mathbf{x}) s(\mathbf{y}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{x}^{(\ell)} \cap \mathbf{y}^{(1)}} \frac{W^{(\ell)}(\mathbf{x}, \mathbf{z}) W^{(1)}(\mathbf{z}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{x}) \nu^{(1)}(\mathbf{z})} \right) \langle s(\mathbf{x}) s(\mathbf{y}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \psi_{\mathbf{xy}}^{(\ell+1)} \langle s(\mathbf{x}) s(\mathbf{y}) \rangle, \end{aligned} \quad (\text{S172})$$



where we have first used Eq. (S132) in going from the second to third line, and then introduced  $W^{(1)}(\mathbf{z}, \mathbf{y}) = 1$  for  $\mathbf{z} \in \mathbf{y}^{(1)}$ , before finally using the definition of  $\psi_{\mathbf{x}, \mathbf{y}}^{(\ell)}$  from Eq. (S169c). In a similar manner one shows that

$$\begin{aligned}
& \frac{1}{N} \sum_{\mathbf{x}} \frac{1}{\nu^{(\ell)}(\mathbf{x})} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{x}, \mathbf{y}) \frac{1}{\nu^{(1)}(\mathbf{x})} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} \langle s(\mathbf{z})s(\mathbf{y}) \rangle \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} \frac{W^{(\ell)}(\mathbf{x}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{x})\nu^{(1)}(\mathbf{x})} \langle s(\mathbf{z})s(\mathbf{y}) \rangle \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{y}^{(\ell)} \cap \mathbf{x}^{(1)}} \frac{W^{(\ell)}(\mathbf{z}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{z})\nu^{(1)}(\mathbf{z})} \right) \langle s(\mathbf{x})s(\mathbf{y}) \rangle \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left( \sum_{\mathbf{z} \in \mathbf{y}^{(\ell)} \cap \mathbf{x}^{(1)}} \frac{W^{(1)}(\mathbf{x}, \mathbf{z})W^{(\ell)}(\mathbf{z}, \mathbf{y})}{\nu^{(\ell)}(\mathbf{z})\nu^{(1)}(\mathbf{z})} \right) \langle s(\mathbf{x})s(\mathbf{y}) \rangle \\
&= \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \phi_{\mathbf{x}\mathbf{y}}^{(\ell+1)} \langle s(\mathbf{x})s(\mathbf{y}) \rangle, \tag{S173}
\end{aligned}$$

where we have first used Eq. (S133) in going from the second to third line, then introduced  $W^{(1)}(\mathbf{x}, \mathbf{z}) = 1$  for  $\mathbf{x} \in \mathbf{z}^{(1)}$ , and finally used the definition of  $\phi_{\mathbf{x}, \mathbf{y}}^{(\ell)}$  from Eq. (S169d). Eq. (S168) then becomes

$$\begin{aligned}
\frac{dC^{(\ell)}}{dt} &= 2\eta \frac{1}{N} \sum_{\mathbf{x}} \Omega_{\mathbf{x}}^{(\ell)} - \alpha \frac{1}{N} \sum_{\mathbf{x}} \left[ 2\Omega_{\mathbf{x}}^{(\ell)} - \chi_{\mathbf{x}}^{(\ell)} \right] S(\mathbf{x}) + \alpha \langle m \rangle - 2\eta C^{(\ell)} \\
&\quad - 2\gamma \frac{1}{N} \sum_{\mathbf{x}} \Omega_{\mathbf{x}}^{(\ell)} S^{(1)}(\mathbf{x}) + \gamma \frac{1}{N} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell+1)}} \left[ \psi_{\mathbf{x}\mathbf{y}}^{(\ell+1)} + \phi_{\mathbf{x}\mathbf{y}}^{(\ell+1)} \right] \langle s(\mathbf{x})s(\mathbf{y}) \rangle. \tag{S174}
\end{aligned}$$

Several of the terms in the last expression take the form of weighted magnetisations, whose steady-state we can determine from Eq. (S159). However the two final terms are problematic and we have not attempted to solve this equation in full generality. Progress is possible for what we will call ‘homogeneous’ networks, as discussed in the next section.

Alternatively we can choose model parameters such that  $\gamma = 0$ . However, this would mean that  $\tau \rightarrow \pm\infty$  and the steady-state configuration would always be random which is not of particular interest (see also Sec. S8).

## E. Homogeneous networks

### 1. Homogeneity assumptions

We define networks that fulfill the following two properties as *homogeneous*:

- (1) the total number walks of length  $\ell$  starting and ending at any node  $\mathbf{x}$  is the same for all  $\mathbf{x}$ ,  $W^{(\ell)}(\mathbf{x}, \mathbf{x}) \equiv W^{(\ell)}$ ;
- (2) the total number of walks of length  $\ell$  starting at  $\mathbf{x}$  (ending at any point) is also the same for all  $\mathbf{x}$ ,  $\nu^{(\ell)}(\mathbf{x}) \equiv \nu^{(\ell)}$ .

Even though the conditions (1) and (2) are fairly strict there are several types of network where they apply. For example, infinite regular lattices in any dimension or Bethe lattices (infinite regular trees) have these properties. Individual realisations of degree-regular networks do not fulfill (1) and (2), but these conditions hold as averages over the ensemble of degree-regular networks, i.e. the expected value of  $W^{(\ell)}(\mathbf{x}, \mathbf{x})$  is the same for all nodes  $\mathbf{x}$ , and similarly for the expected value of  $\nu^{(\ell)}(\mathbf{x})$ .

We now proceed making these assumptions, writing  $W^{(\ell)}$  and  $\nu^{(\ell)}$  for the common values of all  $W^{(\ell)}(\mathbf{x}, \mathbf{x})$  and  $\nu^{(\ell)}(\mathbf{x})$  respectively. The assumptions imply in particular that all nodes in the network have the same degree,  $\nu^{(1)}$ . We note that we then have  $\nu^{(\ell+1)} = \nu^{(\ell)}\nu^{(1)}$ , and from this we find  $\nu^{(\ell)} = (\nu^{(1)})^\ell$ .

Noting again that  $W^{(1)}(\mathbf{z}, \mathbf{y}) = A_{\mathbf{z}\mathbf{y}}$  is equal to one if and only if  $\mathbf{z}$  and  $\mathbf{y}$  are nearest neighbours, and zero otherwise, Eqs. (S169a)-(S169d) can then be simplified as follows:

$$\Omega_{\mathbf{x}}^{(\ell)} = \frac{W^{(\ell)}(\mathbf{x}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{x})} = \frac{W^{(\ell)}}{\nu^{(\ell)}} \equiv \Omega^{(\ell)}, \quad \forall \mathbf{x} \quad (\text{S175a})$$

$$\chi_{\mathbf{x}}^{(\ell)} = \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} \frac{W^{(\ell)}(\mathbf{y}, \mathbf{x})}{\nu^{(\ell)}(\mathbf{y})} = \frac{1}{\nu^{(\ell)}} \sum_{\mathbf{y} \in \mathbf{x}^{(\ell)}} W^{(\ell)}(\mathbf{y}, \mathbf{x}) = 1, \quad \forall \mathbf{x} \quad (\text{S175b})$$

$$\begin{aligned} \psi_{\mathbf{x}\mathbf{y}}^{(\ell+1)} &= \frac{1}{\nu^{(\ell)}(\mathbf{x})} \left[ \sum_{\mathbf{z} \in \mathbf{x}^{(\ell)} \cap \mathbf{y}^{(1)}} W^{(\ell)}(\mathbf{x}, \mathbf{z}) \frac{1}{\nu^{(1)}(\mathbf{z})} W^{(1)}(\mathbf{z}, \mathbf{y}) \right] \\ &= \frac{1}{\nu^{(\ell)} \nu^{(1)}} \sum_{\mathbf{z} \in \mathbf{y}^{(1)}} W^{(\ell)}(\mathbf{x}, \mathbf{z}) W^{(1)}(\mathbf{z}, \mathbf{y}) \\ &= \frac{W^{(\ell+1)}(\mathbf{x}, \mathbf{y})}{\nu^{(\ell+1)}}, \quad \forall \mathbf{x}, \mathbf{y} \end{aligned} \quad (\text{S175c})$$

$$\begin{aligned} \phi_{\mathbf{x}\mathbf{y}}^{(\ell+1)} &= \sum_{\mathbf{z} \in \mathbf{y}^{(\ell)} \cap \mathbf{x}^{(1)}} \left[ \frac{1}{\nu^{(\ell)}(\mathbf{z})} W^{(\ell)}(\mathbf{z}, \mathbf{y}) \frac{1}{\nu^{(1)}(\mathbf{z})} W^{(1)}(\mathbf{z}, \mathbf{x}) \right] \\ &= \frac{1}{\nu^{(\ell+1)}} \sum_{\mathbf{z} \in \mathbf{x}^{(1)}} W^{(\ell)}(\mathbf{z}, \mathbf{y}) \\ &= \frac{W^{(\ell+1)}(\mathbf{x}, \mathbf{y})}{\nu^{(\ell+1)}}. \quad \forall \mathbf{x}, \mathbf{y} \end{aligned} \quad (\text{S175d})$$

## 2. Interface density for homogeneous networks, and resulting simplifications

Assuming that the underlying network is homogeneous (in the sense as defined above), Eq. (S174) reduces to

$$\frac{dC^{(\ell)}}{dt} = 2\Omega^{(\ell)} \left[ \eta - \alpha \langle m \rangle - \gamma C^{(1)} \right] - 2 \left[ \eta C^{(\ell)} - \alpha \langle m \rangle - \gamma C^{(\ell+1)} \right]. \quad (\text{S176})$$

In the steady-state we then have the following relation:

$$\eta C_{\text{st}}^{(\ell)} - \gamma C_{\text{st}}^{(\ell+1)} = \Omega^{(\ell)} \left[ \eta - \alpha \langle m \rangle_{\text{st}} - \gamma C_{\text{st}}^{(1)} \right] + \alpha \langle m \rangle_{\text{st}}. \quad (\text{S177})$$

The interface density, the quantity we are aiming to calculate, is given by

$$\langle \sigma \rangle = \frac{1 - C^{(1)}}{2}, \quad (\text{S178})$$

thus we need to determine  $C_{\text{st}}^{(1)}$ . However setting  $\ell = 0$  in Eq. (S177) is not useful, as all terms involving  $C_{\text{st}}^{(1)}$  then cancel out because of  $\Omega^{(\ell=0)} = 1$ . To see the latter we note that there is one walk of length zero steps starting at  $\mathbf{x}$ , and this walk trivially also ends at  $\mathbf{x}$ . Thus Eq. (S175a) evaluates to one.

Instead, to proceed we multiply both sides of Eq. (S177) by  $\gamma^\ell / \eta^\ell$ , and then sum over  $\ell$  from zero to infinity. We find

$$\sum_{\ell=0}^{\infty} \left( \frac{\gamma^\ell}{\eta^{\ell-1}} C_{\text{st}}^{(\ell)} - \frac{\gamma^{\ell+1}}{\eta^\ell} C_{\text{st}}^{(\ell+1)} \right) - \alpha \langle m \rangle_{\text{st}} \sum_{\ell=0}^{\infty} \frac{\gamma^\ell}{\eta^\ell} = \left[ \eta - \alpha \langle m \rangle_{\text{st}} - \gamma C_{\text{st}}^{(1)} \right] \sum_{\ell=0}^{\infty} \Omega^{(\ell)} \frac{\gamma^\ell}{\eta^\ell}. \quad (\text{S179})$$

The first term on the left-hand side is a telescopic sum and simplifies to

$$\sum_{\ell=0}^{\infty} \left( \frac{\gamma^\ell}{\eta^{\ell-1}} C_{\text{st}}^{(\ell)} - \frac{\gamma^{\ell+1}}{\eta^\ell} C_{\text{st}}^{(\ell+1)} \right) = \eta C_{\text{st}}^{(0)} - \lim_{\ell \rightarrow \infty} \frac{\gamma^{\ell+1}}{\eta^\ell} C_{\text{st}}^{(\ell+1)}. \quad (\text{S180})$$

We can use analogous arguments to that at the end of Sec. S6 B, except considering  $C_{\text{st}}^{(\ell)}$  instead of  $\langle m_n \rangle_{\text{st}}$ , to perform the geometric sum in Eq. (S179) and show that the limit in Eq. (S180) vanishes. Eq. (S179) then becomes

$$\eta C_{\text{st}}^{(0)} - \frac{\alpha \langle m \rangle_{\text{st}}}{1 - \frac{\gamma}{\eta}} = \left[ \eta - \alpha \langle m \rangle_{\text{st}} - \gamma C_{\text{st}}^{(1)} \right] \sum_{\ell=0}^{\infty} \Omega^{(\ell)} \left( \frac{1}{1 + \tau} \right)^{\ell}. \quad (\text{S181})$$

Using the fact that  $C_{\text{st}}^{(0)} = \langle s(\mathbf{x})s(\mathbf{x}) \rangle_{\text{st}} = 1$ , and the identities from Eqs. (S166a) and (S166b), and Eq. (S178), we find

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{HG}}(\tau, \mathbf{A}) \left( 1 - \langle m \rangle_{\text{st}}^2 \right), \quad (\text{S182})$$

with

$$H_{\text{HG}}(\tau) = \frac{1 + \tau}{\sum_{\ell=0}^{\infty} \Omega^{(\ell)} \left( \frac{1}{1 + \tau} \right)^{\ell}} - \tau. \quad (\text{S183})$$

We note that calculation of  $\Omega^{(\ell)}$  requires the adjacency matrix of the network  $\mathbf{A}$ . The subscript ‘HG’ stands for ‘homogeneous graph’, and indicates that the relation holds only when the homogeneity assumptions apply. See Sec. S8 for detailed discussion on  $\tau$  and  $H(\tau)$ .

Eq. (S183) is not closed form but the coefficients  $\Omega^{(\ell)}$  can, in principle, be obtained through direct enumeration of walks on a given network. We note that this only needs to be performed once for a given network, to obtain the function  $H_{\text{HG}}(\tau)$ . We discuss this further in Sec. S6 E 7.

We now move on to use Eq. (S183) for a number of special cases (finite complete networks, hyper-cubic lattices, and Bethe lattices) where closed form solutions can be found.

### 3. First special case: Finite complete networks

We consider a complete network with  $N$  nodes. For any fixed node, we write  $a_N(\ell)$  for the number of walks starting and ending at that node (this number will be the same for all nodes, given the all-to-all connectivity). We note specifically that we consider networks in which the different nodes are distinguishable from one another. Therefore the walk from, say, node 1 to 2 back to 1 is different from the walk from 1 to 3 to 1. One

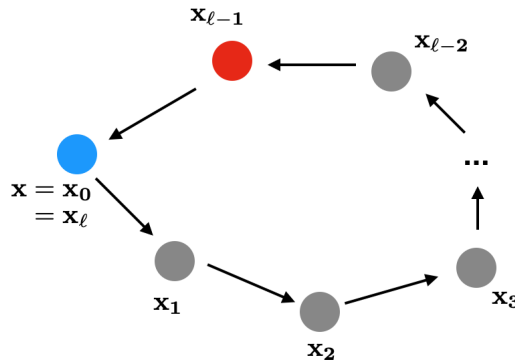


FIG. S6. Illustration of closed walks on a complete network (see text).

then has the following recursion

$$a_N(\ell) = (N - 1)^{\ell-1} - a_N(\ell - 1), \quad (\text{S184})$$

with  $a_N(0) \equiv 1$ .

This recursion can be derived as follows (see Fig. S6), keeping in mind that the walker cannot stay put in any step of the walk. Consider a fixed node  $\mathbf{x}$ , and a closed walk of length  $\ell$  starting and ending at  $\mathbf{x}$ . The walk consists of a sequence of  $\ell + 1$  nodes in total, which we label  $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_\ell$ . The first and the last of these are given by  $\mathbf{x}$  ( $\mathbf{x}_0 = \mathbf{x}_\ell = \mathbf{x}$ ), shown in blue in Fig. S6. Now focus on node  $\mathbf{x}_1$ . This node can be any node in the network other than node  $\mathbf{x}_0$  (the walker must make a hop in the first step). Similarly,  $\mathbf{x}_2$  can be any node other than  $\mathbf{x}_1$ , and so on. If we follow this though until (and including) node  $\mathbf{x}_{\ell-1}$ , we find a total of  $(N - 1)^{\ell-1}$  possible sequences for the nodes  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{\ell-1}$  such that no node is the same as the previous one in the sequence. This produces the term  $(N - 1)^{\ell-1}$  in Eq. (S184). There is however another constraint, and not all of these sequences produce valid walks of length  $\ell$  starting and ending at  $\mathbf{x}$ . In a valid walk, node  $\mathbf{x}_{\ell-1}$  (shown in red in the figure) must not be node  $\mathbf{x}$ , as the walker must make a hop in the  $\ell^{\text{th}}$  step. We therefore need to subtract the number of all sequences for which this is the case. This number is  $a_N(\ell - 1)$ , as the sequence  $\mathbf{x}_0 = \mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1} = \mathbf{x}$  is then a closed walk of length  $\ell - 1$  starting and ending at  $\mathbf{x}$ . This produces the second term in Eq. (S184). Thus, we have demonstrated the validity of the recursion.

One can directly show that (see e.g. sequences A109499, A109500, A109501 in [S10])

$$a_N(\ell) = \frac{(N - 1)^\ell + (N - 1)(-1)^\ell}{N} \quad (\text{S185})$$

fulfills the recursion in Eq. (S184), as well as the condition  $a_N(0) = 1$ .

Given that further the number of all walks of length  $\ell$  starting at  $\mathbf{x}$  on a complete network of size  $N$  is given by  $\nu^{(\ell)}(\mathbf{x}) = (N - 1)^\ell$ , we have

$$\Omega^{(\ell)} = \frac{W^{(\ell)}}{\nu^{(\ell)}(\mathbf{x})} = \frac{a_N(\ell)}{(N - 1)^\ell} = \frac{(N - 1)^\ell + (N - 1)(-1)^\ell}{N(N - 1)^\ell}. \quad (\text{S186})$$

The generating function for the sequence  $\Omega^{(\ell)}$  can be directly evaluated to be

$$\sum_{\ell=0}^{\infty} \Omega^{(\ell)} z^\ell = \frac{1 - \left(\frac{N-2}{N-1}\right) z}{1 - \left(\frac{N-2}{N-1}\right) z - \left(\frac{1}{N-1}\right) z^2}. \quad (\text{S187})$$

Thus Eq. (S182) for the finite complete network is

$$\begin{aligned} \langle \sigma \rangle_{\text{st}} &= \frac{1}{2} \left[ \frac{(N - 1)\tau}{1 + (N - 1)\tau} \right] \left( 1 - \langle m \rangle_{\text{st}}^2 \right) \\ &\equiv \frac{1}{2} H_{\text{CG}}(\tau, N) \left( 1 - \langle m \rangle_{\text{st}}^2 \right), \end{aligned} \quad (\text{S188})$$

(the subscript ‘CG’ stands for ‘complete graph’). We note that in the limit  $N \rightarrow \infty$  we recover Eq. (S29).

#### 4. Second special case: Infinite hyper-cubic lattice in $d$ -dimensions

We next consider the case of a hyper-cubic lattice in  $d$ -dimensions. We then have  $\Omega^{(\ell)} = 0$  for odd values of  $\ell$ , as it is impossible to return to a given point in an odd number of steps on a hyper-cubic lattice. Thus we focus on even  $\ell = 2k$ . So looking at Eq. (S183) we need to calculate  $\Omega^{(\ell)} \rightarrow \Omega_d^{(2k)}$ , where the  $d$  subscript denotes the dimension of the hyper-cubic lattice. From Eq. (S175a) we have

$$\Omega_d^{(2k)} = \frac{W_d^{(2k)}}{\nu_d^{(2k)}}, \quad (\text{S189})$$

On a  $d$ -dimensional hyper-cubic lattice, the total number of walks of length  $2k$  starting at a given point is

$$\nu_d^{(2k)} = (2d)^{2k}, \quad (\text{S190})$$

given that the walker can choose from one of  $2d$  directions at each step. The number of walks starting and ending at the same point is

$$W_d^{(2k)} = \sum_{n_1 + \dots + n_d = k} \frac{(2k)!}{\prod_{i=1}^d (n_i!)^2}. \quad (\text{S191})$$

In this expression the  $n_i$  ( $i = 1, \dots, d$ ) are integer numbers;  $n_i$  indicates how many steps in total the walker takes in the positive  $i$  direction along the walk. Given that the walker returns to the starting point,  $n_i$  is also the number of steps taken in the negative  $i$  direction. Since a total of  $2k$  steps must be made, we have  $\sum_i 2n_i = 2k$ , i.e.  $\sum_i n_i = k$ . The term inside the sum in Eq. (S191) is the number of distinct walks that can be generated for a given choice of  $n_1, \dots, n_d$ . The multinomial coefficient is the number of permutations of the  $n_1$  forward steps in direction 1, the  $n_1$  backward steps in direction 1, the  $n_2$  forward steps in direction 2, the  $n_2$  backward steps in direction 2 and so on, keeping in mind that the total number of steps is  $2k$ .

Using this we can construct the following so-called ‘exponential generating function’ [note the factors of  $1/(2k)!$  which is the difference between an exponential generating function and an ordinary generating function],

$$\sum_{k=0}^{\infty} W_d^{(2k)} \frac{x^{2k}}{(2k)!} = I_0(2x)^d, \quad (\text{S192})$$

where  $I_0(\cdot)$  is the modified Bessel function of the first kind. A proof of this can be found in Sec. S10 A. A Laplace-Borel transform, which we define as

$$\mathcal{L}_B[f](x) = \int_0^{\infty} e^{-t} f(tx) dt, \quad (\text{S193})$$

can be used to convert an exponential generating function of the type in Eq. (S192) to an ordinary generating function [S11, pp. 374, 566], [S12, p. 750] or [S13, p. 2]. Therefore we have

$$\sum_{k=0}^{\infty} W_d^{(2k)} x^{2k} = \int_0^{\infty} e^{-t} [I_0(2xt)]^d dt. \quad (\text{S194})$$

What we want is a summation of the form in Eq. (S183), so we need to incorporate the  $\nu_d^{(2k)} = (2d)^{2k}$  factor which can be done by applying the transform  $x \rightarrow \frac{x}{2d}$  to Eq. (S194),

$$\sum_{k=0}^{\infty} W_d^{(2k)} \frac{x^{2k}}{(2d)^{2k}} = \int_0^{\infty} e^{-t} \left[ I_0\left(\frac{xt}{d}\right) \right]^d dt. \quad (\text{S195})$$

With  $x = \frac{1}{1+\tau}$ , the summation on the LHS is exactly that in Eq. (S183) for an infinite hyper-cubic  $d$ -dimensional lattice. So we are left with solving the integral on the RHS.

To proceed, it is now useful to make the substitution  $xt = z$ ,

$$\begin{aligned} \int_0^{\infty} e^{-t} \left[ I_0\left(\frac{xt}{d}\right) \right]^d dt &= \frac{1}{x} \int_0^{\infty} e^{-\frac{z}{x}} \left[ I_0\left(\frac{z}{d}\right) \right]^d dz \\ &= \frac{1}{x} \mathcal{L} \left( \left[ I_0\left(\frac{z}{d}\right) \right]^d \right). \end{aligned} \quad (\text{S196})$$

where  $\mathcal{L}$  is the standard Laplace transform defined as

$$\mathcal{L}[f](s) = \int_0^{\infty} e^{-sz} f(z) dz, \quad (\text{S197})$$

where in this case  $s = \frac{1}{x}$ . We can now make use of the following general identity [S14, p. 346],

$$\begin{aligned} \mathcal{L} \left( z^\lambda \prod_{i=1}^d I_{\nu_i}(a_i z) \right) &= \frac{\Gamma(\lambda + \nu + 1)}{2^\nu p^{\lambda + \nu + 1}} \left[ \prod_{i=1}^d \frac{a_i^{\nu_i}}{\Gamma(\nu_i + 1)} \right] \\ &\times F_C^{(d)} \left( \frac{\lambda + \nu + 1}{2}, \frac{\lambda + \nu + 2}{2}; \nu_1, \dots, \nu_d + 1; \frac{a_1^2}{p^2}, \dots, \frac{a_d^2}{p^2} \right), \end{aligned} \quad (\text{S198})$$

where  $F_C^{(d)}$  is a Lauricella hypergeometric series [S15, p. 113][S16],

$$F_C^{(d)}(a, b; c_1, \dots, c_d; y_1, \dots, y_d) = \sum_{i_1, \dots, i_d=0}^{\infty} \frac{(a)_{i_1+\dots+i_d} (b)_{i_1+\dots+i_d}}{(c_1)_{i_1} \dots (c_d)_{i_d} i_1! \dots i_d!} y_1^{i_1} \dots y_d^{i_d}, \quad (\text{S199})$$

with

$$(q)_i \equiv q(q+1)\dots(q+i-1), \quad (\text{S200})$$

for  $i = 0, 1, 2, 3, \dots$ , referred to as the Pochhammer symbol or rising factorial.

We note that the identity in Eq. (S198) as printed in [S14] contains a typographical error, for further details see also [S17]. This error has been corrected in Eq. (S198).

We now substitute Eq. (S198) in the special case  $\lambda = 0$ ,  $\nu_i = 0$  and  $a_i = \frac{1}{d}$  into Eq. (S196) and find

$$\int_0^\infty e^{-t} \left[ I_0 \left( \frac{zt}{d} \right) \right]^d dt = F_C^{(d)} \left( \frac{1}{2}, 1; 1, \dots, 1; \frac{1}{[d(1+\tau)]^2}, \dots, \frac{1}{[d(1+\tau)]^2} \right), \quad (\text{S201})$$

where  $z = \frac{1}{1+\tau}$ . With this, we find the following expression for the interface density on a  $d$ -dimensional hyper-cubic lattice,

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} \left[ \frac{1+\tau}{F_C^{(d)} \left( \frac{1}{2}, 1; 1, \dots, 1; \frac{1}{[d(1+\tau)]^2}, \dots, \frac{1}{[d(1+\tau)]^2} \right)} - \tau \right] \left( 1 - \langle m \rangle_{\text{st}}^2 \right). \quad (\text{S202})$$

### 5. Application to infinite two-dimensional square lattice

We have already derived an analytical expression for  $H(\tau)$  for infinite 2D square lattice in Sec. S5 of this Supplement [see Eq. (S120)]. However as a test of our results for homogeneous networks, we can show that Eq. (S183) reduces to Eq. (S120).

We can directly evaluate the summation in Eq. (S183) by first determining a closed form expression for  $\Omega^{(\ell)}$ . On an infinite square lattice only even length walks can return us to the initial position, so it is really  $\Omega^{(2k)}$  we wish to calculate.

The number of closed walks of length  $2k$  can be calculated with the following combinatorial argument. If we take  $p$  steps North, we must then also take  $p$  steps South. There are  $\binom{2k}{2p}$  ways to choose these  $2p$  steps from  $2k$  and  $\binom{2p}{p}$  ways to organise the North and South steps. From the remaining steps, half of them must be East and the other half must be West, there are  $\binom{2k-2p}{k-p}$  ways of arranging these steps. Thus the total number of closed walks of length  $2k$  is

$$\begin{aligned} \sum_{p=0}^k \binom{2k}{2p} \binom{2p}{p} \binom{2(k-p)}{k-p} &= \sum_{p=0}^k \frac{(2k)!}{(2p)!(2(k-p))!} \cdot \frac{(2p)!}{p!^2} \cdot \frac{(2(k-p))!}{(k-p)!^2} \\ &= \frac{(2k)!}{k!^2} \sum_{p=0}^k \frac{k!^2}{p!^2(k-p)!^2} \\ &= \binom{2k}{k} \sum_{p=0}^k \binom{k}{p} \binom{k}{k-p} \\ &= \binom{2k}{k}^2, \end{aligned} \quad (\text{S203})$$

where in the last step we use Vandermonde's identity [S18]:

$$\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}. \quad (\text{S204})$$

We also note that the total number of length  $2k$  walks on a 2D square lattice is  $\nu^{(2k)} = 4^{2k}$ . Thus, letting  $x = \frac{1}{1+\tau}$ , the summation in Eq. (S183) becomes

$$\sum_{k=0}^{\infty} \Omega^{(2k)} x^{2k} = \sum_{k=0}^{\infty} \binom{2k}{k}^2 \frac{x^{2k}}{4^{2k}} = \frac{2}{\pi} K(x), \quad (\text{S205})$$

where we have used the power series for the complete elliptical integral of the first kind [S19, (1.3.6)]. With this, Eq. (S183) reduces exactly to Eq. (S120).

### 6. Third special case: Infinite Bethe lattice

The infinite regular tree, or Bethe lattice, is a network in which every node has degree  $n$  and which has a cascading tree-like structure such that there are no loops or leaf nodes.

On such trees, only walks of even length can return to the starting point. So we have  $W^{(2k+1)} = 0$  for all  $k \in \mathbb{Z}_{\geq 0}$ , and thus only need to perform the summation in Eq. (S183) over even  $k$ . Additionally, it can be seen by inspection that  $\nu^{(\ell)} = n^\ell$ . Ultimately, the summation in Eq. (S183) reduces to

$$T(z) = \sum_{k=0}^{\infty} \mathcal{T}^{(k)} z^k, \quad (\text{S206})$$

for

$$z = \frac{1}{n^2(1+\tau)^2}, \quad (\text{S207})$$

where  $\mathcal{T}^{(k)} \equiv W^{(2k)}$  is the number of self returning walks of length  $2k$  from any point on the network.

This means that we need the generating function for the sequence  $\mathcal{T}^{(k)}$ . To calculate this we follow the approach in [S20] and first consider an infinite rooted tree, in which the root has degree  $r$  and all other nodes have degree  $c+1$ . We write  $\mathbb{T}$  for this tree.

A closed rooted walk is defined as a walk which starts and ends at the root node. We will refer to the nearest neighbours of the root as the ‘children’.

Any closed rooted walk of length  $2k$  in  $\mathbb{T}$  can be broken down as follows:

1. The walker starts at the root, and takes one step from the root to one of its  $r$  nearest neighbours (i.e., a child).
2. A closed walk of length  $2i$ , avoiding the root, is undertaken, starting at the child node in step 1, and bringing the walker back to the same child.
3. The walker returns to the root node, having taken  $2(i+1)$  steps in total.
4. A closed rooted walk of length  $2k - 2(i+1)$  can now be undertaken.

We define  $\mathcal{R}_r^{(k)}$  as the number of self-returning walks of length  $2k$  from the root node, and  $\mathcal{C}_{c+1}^{(k)}$  as the number of self-returning root-avoiding walks of length  $2k$  from one of the children. The subscripts indicate the degree of the root ( $r$ ) and that of all other nodes ( $c+1$ ) respectively. These quantities are specific to the tree  $\mathbb{T}$ . However, the number of self-returning root-avoiding walks starting at a child will not depend on the degree  $r$  of the root.

From the above breakdown on rooted walks of length  $2k$  we can obtain the following recursive relation:

$$\mathcal{R}_r^{(n+1)} = r \sum_{i=0}^n \mathcal{R}_r^{(i)} \mathcal{C}_{c+1}^{(n-i)}. \quad (\text{S208})$$

We next define

$$R_r(z) \equiv \sum_{k=0}^{\infty} \mathcal{R}_r^{(k)} z^k \quad (\text{S209})$$

as the generating function for self-returning walks in the tree from the root node which has degree  $r$ . Similarly,

$$C_{c+1}(z) = \sum_{k=0}^{\infty} \mathcal{C}_{c+1}^{(k)} z^k \quad (\text{S210})$$

is the generating function for self-returning walks in the tree from a child of the root node and avoiding the root.

With these definitions we have

$$\begin{aligned}
R_r(z) &\equiv \sum_{k=0}^{\infty} \mathcal{R}_r^{(k)} z^k \\
&= 1 + \sum_{k=1}^{\infty} \mathcal{R}_r^{(k)} z^k \\
&= 1 + r \sum_{k=1}^{\infty} \sum_{i=0}^{k-1} \mathcal{R}_r^{(i)} \mathcal{C}_{c+1}^{(k-1-i)} z^k \\
&= 1 + rz \sum_{k=0}^{\infty} \sum_{i=0}^k \mathcal{R}_r^{(i)} \mathcal{C}_{c+1}^{(k-i)} z^k \\
&= 1 + rz R_r(z) C_{c+1}(z).
\end{aligned} \tag{S211}$$

The last step can be seen by writing out the first few terms of the series for  $R_r(z)$  and  $C_{c+1}(z)$  in Eqs. (S209) and (S210).

We can determine  $R_r(z)$  from Eq. (S211) if we can calculate  $C_{c+1}(z)$ . To do this, we note again that  $\mathcal{C}_{c+1}^{(k)}$  does not depend on the degree of the root. Hence  $C_{c+1}(z)$  is not a function of the root-degree either. With this in mind, we now consider a tree  $\tilde{\mathbb{T}}$  with root of degree  $c$  and where, as before, all other nodes have degree  $c+1$ . Setting  $r = c$  in Eq. (S211) we then have

$$\tilde{R}_c(z) = 1 + cz \tilde{R}_c(z) \tilde{C}_{c+1}(z), \tag{S212}$$

where the tildes indicate that these are generating functions for the tree  $\tilde{\mathbb{T}}$ . The only difference between the trees  $\mathbb{T}$  and  $\tilde{\mathbb{T}}$  is the degree of the root. Hence  $\tilde{C}_{c+1}(z) = C_{c+1}(z)$ .

Next we realise that  $\tilde{R}_c(z) = \tilde{C}_{c+1}(z)$ . This is because  $\tilde{C}_{c+1}(z)$  is the generating function of self-returning walks starting from a child and avoiding the root on the tree  $\tilde{\mathbb{T}}$ . For the purpose of counting such walks the child effectively has degree  $c$  (the same degree as the root of  $\tilde{\mathbb{T}}$ ). Hence on the tree  $\tilde{\mathbb{T}}$ , the number of self-returning root-avoiding walks of a given length and starting from a child is the same as the number of self-returning walks of this length starting at the root.

Using  $\tilde{R}_c(z) = \tilde{C}_{c+1}(z)$  and  $\tilde{C}_{c+1}(z) = C_{c+1}(z)$  Eq. (S212) thus becomes

$$C_{c+1}(z) = 1 + cz C_{c+1}^2(z). \tag{S213}$$

We can then solve for  $C_{c+1}(z)$  and find the physically relevant solution to be

$$C_{c+1}(z) = \frac{1 - \sqrt{1 - 4zc}}{2zc} \tag{S214}$$

(this is the solution ensuring that  $\lim_{z \rightarrow 0} C_{c+1}(z) = 1$ , i.e. there is exactly one walk of length zero). It is interesting to note that this generating function is closely related to the generating function for the Catalan numbers [S21].

Having now determined  $C_{c+1}(z)$  we can substitute this back into Eq. (S211) to determine  $R_r(z)$ ,

$$R_r(z) = \frac{2c}{2c - r + r\sqrt{1 - 4zc}}. \tag{S215}$$

For the special case of the  $n$ -regular tree (i.e. the tree where all nodes have the same degree  $n$ ) we simply set  $n = r = c+1$  in Eq. (S215). This gives the generating function for self-returning walks on an infinite  $n$ -regular tree, which we define as  $T(z)$  in Eq. (S206),

$$T(z) = \frac{n-1}{n-2 + n\sqrt{1 - 4(n-1)z}}. \tag{S216}$$



Then with  $z$  as in Eq. (S207) we have a closed-form expression for the summation in Eq. (S206):

$$T(z) = \frac{2(n-1)}{n-2+n\sqrt{1-4(n-1)\left(\frac{1}{n(1+\tau)}\right)^2}}. \quad (\text{S217})$$

With the replacement  $n \rightarrow \mu$ , Eqs. (S182) and (S217) combine to give

$$\begin{aligned} \langle \sigma \rangle_{\text{st}} &= \frac{1}{2} \left[ \frac{\mu-2-\mu\tau + \sqrt{(\tau+1)^2\mu^2 - 4(\mu-1)}}{2(\mu-1)} \right] \left( 1 - \langle m \rangle_{\text{st}}^2 \right) \\ &\equiv \frac{1}{2} H_{\text{BL}}(\tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right), \end{aligned} \quad (\text{S218})$$

(where ‘BL’ stands for ‘Bethe lattice’). We note that this is the same result that we derived in Sec. S3 using the pair approximation for infinite uncorrelated networks. In the current section we did not have to make any approximations about spin correlations as we did for the pair approximation. The only assumption that we made was that the dynamics was being taking place on a particular topology, hence this result is exact for Bethe lattices. This is not surprising as the pair approximation is known to be exact on Bethe lattices [S22, p; 346] as there are no loops.

### 7. Calculation of $\Omega^{(\ell)}$ and validity of the method

For networks that are not homogeneous in the above sense, we can still apply the approach based on network walks [Eq. (S183)] as an approximation. We replace  $\Omega^{(\ell)}$  with its average over all nodes,

$$\Omega^{(\ell)} \rightarrow \frac{1}{N} \sum_{\mathbf{x}} \Omega_{\mathbf{x}}^{(\ell)} \equiv \overline{\Omega^{(\ell)}}. \quad (\text{S219})$$

It is difficult to find this object in closed form. However, for many networks we find that  $\overline{\Omega^{(\ell)}} \rightarrow \frac{1}{N}$  for large  $\ell$  (where  $N$  is the size of the network). Thus we compute  $\overline{\Omega^{(\ell)}}$  directly from the adjacency matrix for  $\ell$  below some cutoff  $L$ , and use  $\overline{\Omega^{(\ell)}} \approx \frac{1}{N}$  for  $\ell > L$ . This approximation should be tested for each network on an individual basis, from which an appropriate choice for  $L$  should be made. Based on this then we can write, using the geometric series,

$$\sum_{\ell=0}^{\infty} \overline{\Omega^{(\ell)}} x^{\ell} \approx \frac{1}{N} \frac{x^{L+1}}{1-x} + \sum_{\ell=0}^L \overline{\Omega^{(\ell)}} x^{\ell}. \quad (\text{S220})$$

In Fig. S7 we show the convergence of  $\overline{\Omega^{(\ell)}}$  to  $\frac{1}{N}$  for Barabási–Albert networks of different size. These networks are not homogeneous. In this case it would be suitable to take  $L \approx 30$ .

Although the approaches based on network walks is only exact on homogeneous networks, we find that it serves as a satisfactory approximation for networks in which the variation of  $\Omega_{\mathbf{x}}^{\ell}$  across nodes is small. In Fig. S8 we show the frequencies at which length  $\ell$  walks (top) and length  $\ell$  loops (bottom) appear as we vary  $\ell$  for degree-regular, Erdős–Rényi, and Barabási–Albert networks. To clarify, for a chosen  $\ell$ , i.e. a particular row of Fig. S8, the bars in each panel report the number of nodes from which it is possible to perform a specific number of length- $\ell$  random walks. For example, choosing  $\ell = 2$ , for a degree-regular network with  $\mu = 8$  [Fig. S8(a1)], all nodes have exactly 64 walks of length  $\ell = 2$ . Moving to an Erdős–Rényi network, Fig. S8(a2), we have a more of a distribution. In this case for example, we see that the majority of nodes have  $\sim 50$  length  $\ell = 2$  random walks starting from that node and ending anywhere. But much fewer nodes have  $\sim 150$  length  $\ell = 2$  random walks. The lower histogram, Fig S8(d1)-(f3), conveys the same thing except the random walks must return to the starting node.

On degree-regular networks with degree  $k$ , the number of walks of a given length  $\ell$  starting from any node is always  $k^{\ell}$ , as confirmed in Fig. S8(a1), (a2) and (a3). The number of loops can vary slightly from node to node (see panels d1-f1) as the network is not exactly homogeneous, but the distributions are still relatively tight, compared to the two other types of network shown.

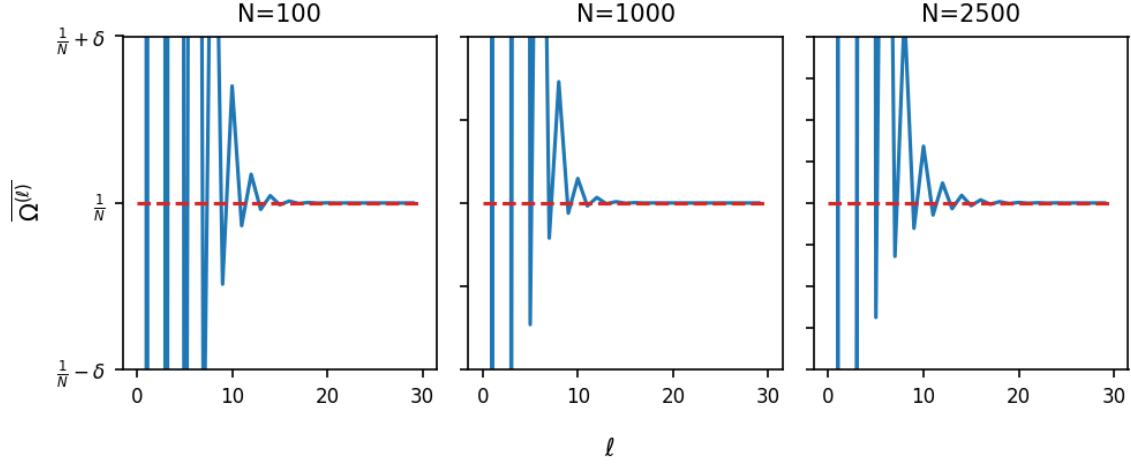


FIG. S7. Demonstration of  $\overline{\Omega^{(\ell)}}$  converging to  $\frac{1}{N}$  for increasing  $\ell$  for different size Barabási–Albert networks with 4 new links established upon adding a new node. The extent of the vertical axis shown is from  $1/N - \delta$  to  $1/N + \delta$ , with  $\delta = 1 \times 10^{-4}$ .

For Erdős–Rényi networks there is more variance in the distributions, and for Barabási–Albert networks there is significantly more. Thus we can anticipate the approximation to be somewhat accurate on Erdős–Rényi networks, and less so on Barabási–Albert networks.

In Fig. S9 we show how  $H_{\text{HG}}(\tau)$  from Eq. (S183) varies as a function of  $\tau$  for different degree-regular (left), Erdős–Rényi (middle) and Barabási–Albert (right) networks. The theory works very well for degree-regular networks and is a good approximation for Erdős–Rényi networks as expected because the variance of their degree distributions is small. For Barabási–Albert there is greater deviation.

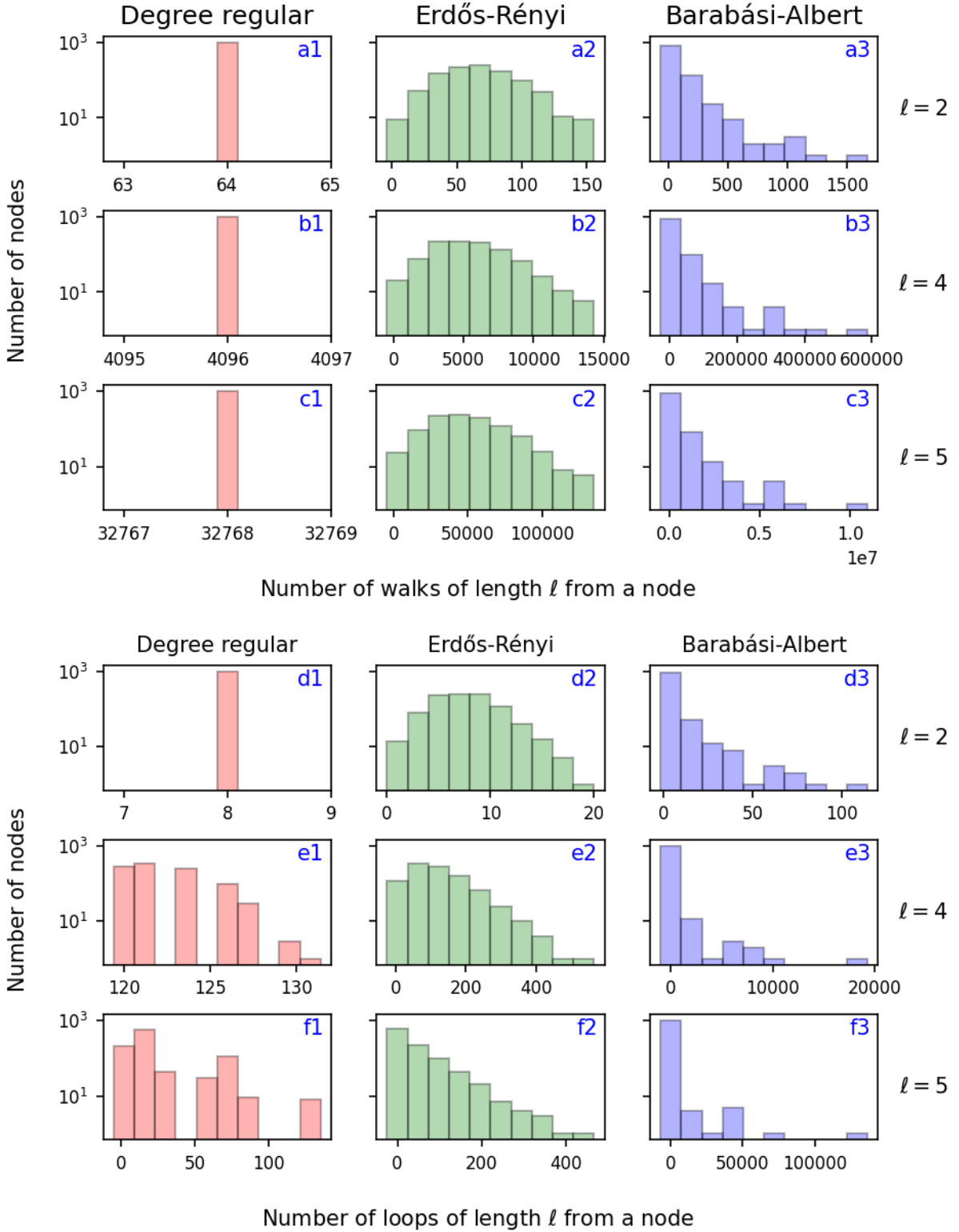


FIG. S8. Histograms of the frequency of length  $\ell$  walks (top) and loops (bottom) from each node in three different networks: degree-regular (left column) with  $\mu = 8$ , Erdős-Rényi (middle column) with  $p = 8/N$  and Barabási-Albert (right column) with 4 new links established upon adding a new node. All networks have  $N = 1,000$ .

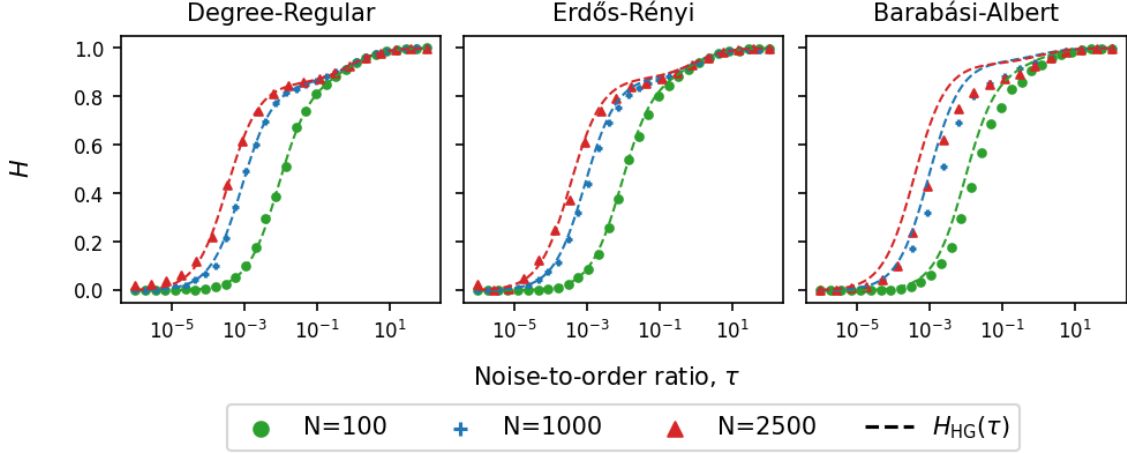


FIG. S9. Plots of  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)]$ , which characterises the scattering, as a function of the noise-to-order ratio,  $\tau$ , for networks of size  $N$ . Degree-regular with  $\mu = 8$  (left), Erdős-Rényi with  $p = 8/N$  (middle), Barabási-Albert with 4 new links established upon adding a new node (right). For all three types of network we have  $\mu \approx 8$  (deviations are due to the finite size of the networks). The dashed lines are the results from our theory, Eq. (S183). Each marker is the result from averaging 1,000 independent simulations in the steady-state at a particular  $\tau$  value. Model parameters for a particular  $\tau$  can be generated via the algorithm detailed in Sec. S10 B.

## S7. Stochastic pair approximation

The stochastic pair approximation (SPA) is an extension of the regular pair approximation discussed in Sec. S3. The SPA targets large but finite uncorrelated networks. This was originally proposed in [S23] (pp 7-16 in particular), which studied the noisy voter model. Our model is more general, and reduces to the noisy voter model in the limit of no copying errors in the horizontal process,  $h'_+ = h'_- = 0$ , and when the vertical process is symmetric,  $v_+ = v_-$ . We here construct the SPA for the more general setup of our model. More precisely we develop what is called ‘S2PA’ in [S23] but here we refer to it as simply ‘SPA’.

Unlike in the main paper and in the remaining sections of this supplement we do not use  $-1$  and  $+1$  for the node states. Instead we denote these states by  $0$  and  $1$ . This is to keep the notation as close to that of [S23].

### A. Master equation for reduced set of degrees of freedom

To begin we consider the reduced set of variables  $\{\mathbf{N}_1, \mathcal{L}\}$  where  $\mathbf{N}_1$  is the vector

$$\mathbf{N}_1 = (N_{1,k_{\min}}, \dots, N_{1,k_{\max}}), \quad (\text{S221})$$

whose components  $N_{1,k}$  are the number of spin-up nodes with degree  $k$ . The notation  $N_{0,k}$  will be used to represent the number of spin-down nodes with degree  $k$ , and thus the total number of nodes of degree  $k$  is  $N_k = N_{0,k} + N_{1,k}$ .  $\mathcal{L}$  is the number of active links. We also define  $n_i \in [0, k_i]$  as the number of neighbours of node  $i$  in the up-state.

When a node  $i$  flips from the down to the up-state, we have the process  $N_{1,k_i} \rightarrow N_{1,k_i} + 1$ . The number of active links connected to  $i$  before flipping is  $n_i$ , after the flip it will be  $k_i - n_i$ , thus a change of  $k_i - 2n_i$  in the number of active links connected to  $i$ . Analogously, when a node flips from up to down, we have  $N_{1,k_i} \rightarrow N_{1,k_i} - 1$  and the total number of active links changes as  $\mathcal{L} \rightarrow \mathcal{L} - k_i + 2n_i$ .

We then define the probability distribution of  $\{\mathbf{N}_1, \mathcal{L}\}$  at time  $t$  as  $P(\mathbf{N}_1, \mathcal{L}; t)$  and we can construct the

following master equation:

$$\begin{aligned} \frac{\partial}{\partial t} P(\mathbf{N}_1, \mathcal{L}; t) = \sum_k \sum_{n=0}^k \left\{ (E_{N_{1,k}}^+ E_{\mathcal{L}}^{k-2n} - 1) [W^{(-,k,n)} P(\mathbf{N}_1, \mathcal{L}; t)] \right. \\ \left. + (E_{N_{1,k}}^- E_{\mathcal{L}}^{-k+2n} - 1) [W^{(+,k,n)} P(\mathbf{N}_1, \mathcal{L}; t)] \right\}, \end{aligned} \quad (\text{S222})$$

where  $W^{(+,k,n)}$  and  $W^{(-,k,n)}$  are the effective rates of the two processes described above. The  $E$ 's are step operators which act on functions of  $N_{1,k}$  or  $\mathcal{L}$  as follows

$$E_{N_{1,k}}^+ f(N_{1,k}) \rightarrow f(N_{1,k} + 1), \quad (\text{S223a})$$

$$E_{\mathcal{L}}^{k-2n} f(\mathcal{L}) \rightarrow f(\mathcal{L} + k - 2n), \quad (\text{S223b})$$

$$E_{N_{1,k}}^- f(N_{1,k}) \rightarrow f(N_{1,k} - 1), \quad (\text{S223c})$$

$$E_{\mathcal{L}}^{-k+2n} f(\mathcal{L}) \rightarrow f(\mathcal{L} - k + 2n). \quad (\text{S223d})$$

Next we define the degree specific magnetisation  $m_k$ , i.e. the magnetisation among the nodes of degree  $k$ . The interface density is denoted by  $\sigma$ . These two quantities can be written in terms of the variables  $\{\mathbf{N}_1, \mathcal{L}\}$  as follows:

$$m_k = 2 \frac{N_{1,k}}{N_k} - 1, \quad (\text{S224a})$$

$$\sigma = \frac{\mathcal{L}}{\frac{1}{2} \mu N}, \quad (\text{S224b})$$

where  $\mu$  is the mean degree of the network, and  $N$  is the total number of nodes, thus  $\frac{1}{2} \mu N$  gives the total number of links.

## B. Dynamics of first and second moments

Eq. (S222) can be used to derive differential equations for the average degree specific magnetisation,  $\langle m_k(t) \rangle$ , the average density of active links,  $\langle \sigma(t) \rangle$ , and correlations such as  $\langle m_k(t) m_{k'}(t) \rangle$ . Angle brackets, as always, represent averages over independent realisations of the dynamics i.e.

$$\langle F[\mathbf{s}(t)] \rangle = \sum_{\mathbf{s}} F[\mathbf{s}(t)] P(\mathbf{s}; t), \quad (\text{S225})$$

where the sum is over all  $2^N$  spin configurations  $\mathbf{s} = (s_1, \dots, s_N)$ , and where  $P(\mathbf{s}; t)$  is the probability of finding the spins in the configuration  $\mathbf{s}$  at time  $t$ . Specifically one finds (after some algebra),

$$\frac{d \langle m_k \rangle}{dt} = \langle F_k \rangle, \quad (\text{S226a})$$

$$\frac{d \langle \sigma \rangle}{dt} = \langle F_\sigma \rangle, \quad (\text{S226b})$$

$$\frac{d \langle m_k m_{k'} \rangle}{dt} = \langle m_{k'} F_k \rangle + \langle m_k F_{k'} \rangle + \delta_{k,k'} \frac{\langle G_k \rangle}{N}, \quad (\text{S226c})$$

where

$$F_k = \frac{2}{N_k} \sum_{n=0}^k [W^{(+,k,n)} - W^{(-,k,n)}], \quad (\text{S227a})$$

$$F_\sigma = \frac{2}{\mu N} \sum_k \sum_{n=0}^k (k - 2n) [W^{(+,k,n)} - W^{(-,k,n)}], \quad (\text{S227b})$$

$$G_k = \frac{4}{P_k N_k} \sum_{n=0}^k [W^{(+,k,n)} + W^{(-,k,n)}], \quad (\text{S227c})$$

and where we have written  $P_k = \frac{N_k}{N}$  for the degree distribution of the network.

We now proceed to calculate the effective rates.  $W^{(+,k,n)}$  is the product of the number of degree  $k$  nodes in the down-state, the probability of a node in the down-state with degree  $k$  having  $n$  neighbours in the up-state,  $P_0(k, n)$ , and the rate at which such a down-state node flips to an up-state node,  $R_{k,n}^+$ . The latter is determined by the dynamics of the model. Therefore,

$$W^{(+,k,n)} = (N_k - N_{1,k})P_0(k, n)R_{k,n}^+. \quad (\text{S228})$$

In a similar manner,

$$W^{(-,k,n)} = N_{1,k}P_1(k, n)R_{k,n}^-, \quad (\text{S229})$$

where  $P_1(k, n)$  is the probability of a node in the up-state with degree  $k$  having  $n$  neighbours in the up-state, and  $R_{k,n}^-$  is the rate at which such a node in the up-state flips to a down-state node.

The spin flip rates in our model are different from the ones in [S23]. We have

$$R_{k,n}^+ = v'_+ + h_+ \frac{n}{k} + h'_+ \frac{k-n}{k}, \quad (\text{S230a})$$

$$R_{k,n}^- = v'_- + h_- \frac{k-n}{k} + h'_- \frac{n}{k}. \quad (\text{S230b})$$

The probabilities  $P_0(k, n)$  and  $P_1(k, n)$  are intricate objects, determined by the distribution of the full spin configuration on the network. In order to make analytical progress we use the pair approximation and describe these probabilities in terms of  $\{\mathbf{N}_1, \mathcal{L}\}$ . This is the same procedure as in Sec. S3, specifically Eq. (S34), where we assumed that the probability of a degree  $k$  node having  $n$  neighbours in the opposite state is binomial. Using this approximation we have

$$P_0(k, n) = \binom{k}{n} c_0^n (1 - c_0)^{k-n}, \quad (\text{S231a})$$

$$P_1(k, n) = \binom{k}{n} c_1^{k-n} (1 - c_1)^n. \quad (\text{S231b})$$

In these expressions  $c_{0/1}$  is the probability that a given neighbour of a node in the down/up-state is in the up/down-state (i.e. in both cases, the probability that a link, connected to the focal node, is active). Formally,

$$c_{0/1} = \frac{\mathcal{L}}{\sum_k k N_{0/1,k}}, \quad (\text{S232})$$

which is the ratio between the total number of active links and the number of links connected emanating from nodes in the up/down-state respectively. Evaluating the above expression gives

$$c_{0/1} = \frac{\sigma}{1 \mp m_L}, \quad (\text{S233})$$

where  $m_L$  is the degree averaged magnetisation defined by

$$m_L = \frac{1}{\mu} \sum_k k P_k m_k. \quad (\text{S234})$$

With these ingredients we can now evaluate the objects on the right-hand sides of Eqs. (S227a)-(S227c) further. Due to the complexity this is done using Mathematica. An accompanying Mathematica notebook can be found at the following GitHub repository [S24].

The resulting expressions for the objects in Eqs. (S227a)-(S227c) are long, so we do not report them here in full. We include one expression as an example,

$$F_k = \alpha - \beta m_k + \frac{\sigma}{1 - m_L^2} \left[ 2\gamma(m_L - m_k) + (\gamma_+ - \gamma_-)(1 - m_k m_L) \right], \quad (\text{S235})$$

where we have used the shorthands from Eq. (S4).

As before we now make the restriction from Eq. (S5),  $\gamma_+ = \gamma_-$ , which helps maintain analytic tractability. The different quantities in Eqs. (S227a)-(S227c) can be expressed in terms of  $m_k$ ,  $m_L$ ,  $m$ ,  $\mu$ ,  $\mu_2$ , and the shorthand parameter combinations defined in Eq. (S4). The quantity  $m$  is the global magnetisation used throughout this paper, and can be calculated from the degree specific magnetisation  $m_k$ ,

$$m = \sum_k P_k m_k. \quad (\text{S236})$$

We use the notation

$$\mu_n = \sum_k P_k k^n \quad (\text{S237})$$

for the  $n^{\text{th}}$  moment of the degree distribution.

We also point out a typo in the second moment of  $P_1(k, n)$  given on p. 9 of [S23]. This should in fact be  $\sum_{n=0}^k n^2 P_1(k, n) = k [k - (2k - 1)c_1 + (k - 1)c_1^2]$ .

We find that Eqs. (S226a)-(S226c) are not closed. However we can use Eq. (S226a) to obtain a differential equation for  $\langle m_L \rangle$ . To do this, we evaluate  $\frac{1}{\mu} \sum_k k P_k \langle F_k \rangle$ , and find

$$\frac{d \langle m_L \rangle}{dt} = \alpha - \beta \langle m_L \rangle. \quad (\text{S238})$$

As a consequence we have

$$\langle m_L \rangle = \langle m_L \rangle_{\text{st}} - [\langle m_L \rangle_{\text{st}} - \langle m_L(0) \rangle] e^{-\beta t}, \quad (\text{S239})$$

and the steady-state value of  $\langle m_L \rangle$  is

$$\langle m_L \rangle_{\text{st}} = \frac{\alpha}{\beta} = \frac{(h'_+ - h'_-) + (v'_+ - v'_-)}{(h'_+ + h'_-) + (v'_+ + v'_-)}. \quad (\text{S240})$$

This is the same solution that we have found in all other cases, for instance see Sec. S3 B where we analysed the standard pair approximation. This is not surprising, as all we have done in this section is perform the pair approximation again but in a style better suited to the extension we will make in Sec. S7 C. In fact, we have already shown in Sec. S6 C that Eq. (S240) is a very general solution, a fact we will use in the next section.

### C. System-size expansion

We now use what is called ‘S2PA’ (stochastic pair approximation 2) in [S23] to address Eqs. (S226a)-(S226c). This method focuses on expanding around the deterministic solution for the variable  $m_L$ .

Motivated by the fact that  $\langle m_L \rangle$  is a self governed variable, as we saw from Eq. (S238), and that the steady-state interface density under the standard pair approximation is as in Eq. (S43), we purpose the following expansion:

$$m_k = m_L + \frac{\lambda_k}{\sqrt{N}} + \frac{\nu_k}{N} + \dots, \quad (\text{S241a})$$

$$\sigma = \frac{1}{2} H_{\text{PA}}(\tau) (1 - m_L^2) + \frac{\lambda_\sigma}{\sqrt{N}} + \dots, \quad (\text{S241b})$$

where  $\lambda_k$ ,  $\nu_k$  and  $\lambda_\sigma$  are stochastic variables. All objects in these equations, excluding  $H_{\text{PA}}(\tau)$  which is given in Eq. (S45), are understood to be time-dependent. Ultimately however, we are only interested in the stationary state values. We neglect higher-order terms in the expansion in powers of  $N^{-1/2}$ . The goal is to calculate  $\langle \sigma \rangle_{\text{st}}$  to leading and sub-leading order. Any expansions we perform in the remainder of this section can be found in a Mathematica notebook at the following GitHub repository [S24].

This expansion is further justified by the fact that individual simulated trajectories in  $(m, \sigma)$ -space, performed on finite size uncorrelated networks, fluctuate around a parabola with height  $H_{\text{PA}}(\tau)$ . This is demonstrated in Fig. 2 in the main paper. Simulated trajectories on a truly infinite uncorrelated network would move exactly along this parabola, in fact they would follow the path described by Eqs. (S25) and (S42) which is also shown in Fig. 2. Time averaging such analytical trajectories on infinite networks in the steady-state (equivalently we can average over many such independent trajectories at a single time in the steady state) will result in the point given by Eqs. (S240) and (S43). However, stochasticity means that the steady-state averages in finite networks sit on a different parabola. This is what we wish to determine and what motivates the system-size expansion in Eqs. (S241a) and (S241b).

To continue, we note that based on the results from Sec. S6 C, that any normalised weighted magnetisation in the steady-state is equal to the steady-state global magnetisation, Eq. (S159). The degree-specific average magnetisation  $\langle m_k \rangle_{\text{st}}$  is effectively a weighted magnetisation with weights  $\delta_{k,k_i}$ , i.e. we have

$$\langle m_k \rangle = \frac{\sum_i \delta_{k,k_i} S_i}{\sum_i \delta_{k,k_i}}, \quad (\text{S242})$$

where  $S_i$  is the average spin state at site  $i$ . Thus we have  $\langle m_k \rangle_{\text{st}} = \langle m \rangle_{\text{st}}$ . Given the definition of  $m_L$  in Eq. (S234) this then also means that  $\langle m_k \rangle_{\text{st}} = \langle m_L \rangle_{\text{st}}$  for all  $k$ . With this we see immediately from Eq. (S241a) that  $\langle \lambda_k \rangle_{\text{st}} = 0$  and  $\langle \nu_k \rangle_{\text{st}} = 0$ .

Next we substitute Eqs. (S241a) and (S241b) into Eqs. (S226b) and (S226c), and equate powers of  $N^{-1/2}$ . Many model parameters that appear, i.e.  $\alpha, \beta, \gamma$  etc. from Eq. (S4), can be expressed in terms of the noise-to-order ratio  $\tau = \frac{\beta}{\gamma}$  from Eq. (S46) and the steady-state global magnetisation  $\langle m \rangle_{\text{st}} = \frac{\alpha}{\beta}$  from Eq. (S24). We find

$$\begin{aligned} \frac{d\langle \lambda_\sigma \rangle}{dt} = & -2 \left\{ \tau - 1 + 2H_{\text{PA}}(\tau) + \frac{2}{\mu} [1 - H_{\text{PA}}(\tau)] \right\} \langle \lambda_\sigma \rangle \\ & + \frac{2\gamma}{\mu} H_{\text{PA}}(\tau) [1 - H_{\text{PA}}(\tau)] \sum_k P_k \langle \lambda_k m_L \rangle \end{aligned} \quad (\text{S243})$$

and

$$\frac{d\langle \lambda_k m_L \rangle}{dt} = -\gamma [H_{\text{PA}}(\tau) + 2\tau] \langle \lambda_k m_L \rangle + \gamma \tau \langle m \rangle_{\text{st}} \langle \lambda_k \rangle \quad (\text{S244})$$

respectively. We note that to get Eq. (S244) we set  $k = k'$  in Eq. (S226c).

Taking both of these equations to the steady-state, and given that we already know that  $\langle \lambda_k \rangle_{\text{st}} = 0$ , we also find that  $\langle \lambda_\sigma \rangle_{\text{st}} = 0$ . With this  $\langle \sigma \rangle_{\text{st}}$  reduces to

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{PA}}(\tau) \left( 1 - \langle m_L^2 \rangle_{\text{st}} \right), \quad (\text{S245})$$

and we are left with determining  $\langle m_L^2 \rangle_{\text{st}}$ . Comparing with the result from the conventional pair approximation in Eq. (S43), we note that the final result contains the object  $\langle m_L^2 \rangle_{\text{st}}$  instead of  $\langle m \rangle_{\text{st}}^2$ . In particular the square is now inside the average.

To calculate  $\langle m_L^2 \rangle_{\text{st}}$ , we start from Eq. (S226c),

$$\frac{d\langle m_k m_{k'} \rangle}{dt} = \langle m_{k'} F_k \rangle + \langle m_k F_{k'} \rangle + \delta_{k,k'} \frac{\langle G_k \rangle}{N}. \quad (\text{S246})$$

We expand this equation in powers of  $\frac{1}{\sqrt{N}}$  using Eqs. (S241a). The LHS of Eq. (S246) then becomes

$$\begin{aligned} \frac{d\langle m_k m_{k'} \rangle}{dt} = & \frac{d\langle m_L^2 \rangle}{dt} + \frac{1}{\sqrt{N}} \left( \frac{d\langle \lambda_k m_L \rangle}{dt} + \frac{d\langle \lambda_{k'} m_L \rangle}{dt} \right) \\ & + \frac{1}{N} \left( \frac{d\langle \lambda_k \lambda_{k'} \rangle}{dt} + \frac{\langle \nu_k m_L \rangle}{dt} + \frac{\langle \nu_{k'} m_L \rangle}{dt} \right) + \mathcal{O}(N^{-3/2}). \end{aligned} \quad (\text{S247})$$



Next we consider the term  $\langle m_{k'} F_k \rangle$  on the RHS of Eq. (S246). This expands as [S24]

$$\begin{aligned} \langle m_{k'} F_k \rangle &= \gamma\tau [\langle m \rangle_{\text{st}} \langle m_L \rangle - \langle m_L^2 \rangle] \\ &\quad + \frac{\gamma}{\sqrt{N}} \left[ - \left( H_{\text{PA}}(\tau) + \tau \right) \langle \lambda_k m_L \rangle - \tau \langle \lambda_{k'} m_L \rangle + \tau \langle \lambda_{k'} \rangle \langle m \rangle_{\text{st}} \right] \\ &\quad + \frac{\gamma}{N} \left[ - \left( H_{\text{PA}}(\tau) + \tau \right) \langle \lambda_k \lambda_{k'} \rangle + \tau \langle \nu_k \rangle \langle m \rangle_{\text{st}} - 2 \left\langle \frac{m_L \lambda_\sigma \lambda_{k'}}{1 - m_L^2} \right\rangle \right. \\ &\quad \left. - \tau \langle \nu_k m_L \rangle - \left( H_{\text{PA}}(\tau) + \tau \right) \langle m_L \nu_{k'} \rangle \right] \\ &\quad + \mathcal{O}(N^{-3/2}). \end{aligned} \quad (\text{S248})$$

The  $\langle m_k F_{k'} \rangle$  term gives a similar expression, but with  $k$  and  $k'$  interchanged.

Since  $\langle G_k \rangle$  carries a pre-factor  $1/N$  in Eq. (S246) we only consider the leading order term of  $\langle G_k \rangle$ . We have [S24]

$$\langle G_k \rangle = \frac{2\gamma}{P_k} \left[ H_{\text{PA}}(\tau) + \tau - H_{\text{PA}}(\tau) \langle m_L^2 \rangle - \tau \langle m \rangle_{\text{st}} \langle m_L \rangle \right] + \mathcal{O}(N^{-1/2}). \quad (\text{S249})$$

Collecting terms in the expansions of the left-hand and right-hand sides of Eq. (S246), and keeping only expressions up to and including order  $1/N$ , we have

$$\begin{aligned} \frac{d \langle m_L^2 \rangle}{dt} + \frac{1}{\sqrt{N}} \left( \frac{d \langle \lambda_k m_L \rangle}{dt} + \frac{d \langle \lambda_{k'} m_L \rangle}{dt} \right) + \frac{1}{N} \left( \frac{d \langle \lambda_k \lambda_{k'} \rangle}{dt} + \frac{\langle \nu_k m_L \rangle}{dt} + \frac{\langle \nu_{k'} m_L \rangle}{dt} \right) = \\ 2\gamma\tau [\langle m \rangle_{\text{st}} \langle m_L \rangle - \langle m_L^2 \rangle] \\ + \frac{\gamma}{\sqrt{N}} \left\{ - \left( H_{\text{PA}}(\tau) + 2\tau \right) \left[ \langle \lambda_k m_L \rangle + \langle \lambda_{k'} m_L \rangle \right] + \tau \langle m \rangle_{\text{st}} \left[ \langle \lambda_k \rangle + \langle \lambda_{k'} \rangle \right] \right\} \\ + \frac{\gamma}{N} \left\{ - 2 \left( H_{\text{PA}}(\tau) + \tau \right) \langle \lambda_k \lambda_{k'} \rangle + \tau \langle m \rangle_{\text{st}} \left[ \langle \nu_k \rangle + \langle \nu_{k'} \rangle \right] - 2 \left\langle \frac{(\lambda_k + \lambda_{k'}) m_L \lambda_\sigma}{1 - m_L^2} \right\rangle \right. \\ \left. - \left( H_{\text{PA}}(\tau) + 2\tau \right) \left[ \langle \nu_k m_L \rangle + \langle \nu_{k'} m_L \rangle \right] \right. \\ \left. + \frac{2}{P_k} \left[ H_{\text{PA}}(\tau) + \tau - H_{\text{PA}}(\tau) \langle m_L^2 \rangle - \tau \langle m \rangle_{\text{st}} \langle m_L \rangle \right] \delta_{k,k'} \right\}. \end{aligned} \quad (\text{S250})$$

Performing the operation  $\sum_{kk'} kk' P_k P_{k'}$  on both sides, several of the contributions can be eliminated. To see this, we perform the sum  $\sum_k k P_k$  on the expansion of  $m_k$  from Eq. (S241a), and find

$$\sum_k k P_k m_k = \sum_k k P_k m_L + \frac{1}{\sqrt{N}} \sum_k k P_k \lambda_k + \frac{1}{N} \sum_k k P_k \nu_k, \quad (\text{S251})$$

resulting in

$$\mu m_L = \mu m_L + \frac{1}{\sqrt{N}} \sum_k k P_k \lambda_k + \frac{1}{N} \sum_k k P_k \nu_k. \quad (\text{S252})$$

Equating powers of  $N^{-1/2}$  we therefore conclude

$$\sum_k k P_k \lambda_k = 0, \quad (\text{S253a})$$

$$\sum_k k P_k \nu_k = 0. \quad (\text{S253b})$$

We now use the identities in Eqs. (S253a) and (S253b) when performing the operation  $\sum_{kk'} kk' P_k P_{k'}$  in Eq. (S250). This results in all stochastic terms vanishing and we are left with

$$\begin{aligned} \frac{d \langle m_L^2 \rangle}{dt} &= -2\gamma\tau \left[ \langle m_L^2 \rangle - \langle m \rangle_{\text{st}} \langle m_L \rangle \right] \\ &\quad + \frac{2\gamma\mu_2}{N\mu^2} \left[ H_{\text{PA}}(\tau) + \tau - H_{\text{PA}}(\tau) \langle m_L^2 \rangle - \tau \langle m \rangle_{\text{st}} \langle m_L \rangle \right], \end{aligned} \quad (\text{S254})$$

where we note the introduction of  $\mu_2$  and  $\mu^2$  arising from terms of the form  $\sum_k k^2 P_k$  and  $(\sum_k k P_k)^2$ . It is possible to solve this equation analytically and obtain a full time dependent solution for  $\langle m_L^2(t) \rangle$  and thus  $\langle \sigma(t) \rangle$  using Eq. (S239).

However, we are mostly interested in the steady-state, so we move to that regime. Eq. (S254) then becomes

$$0 = -\tau N \left[ \langle m_L^2 \rangle_{\text{st}} - \langle m \rangle_{\text{st}}^2 \right] + \frac{\mu_2}{\mu^2} \left[ H_{\text{PA}}(\tau) + \tau - H_{\text{PA}}(\tau) \langle m_L^2 \rangle_{\text{st}} - \tau \langle m \rangle_{\text{st}}^2 \right] \quad (\text{S255})$$

and we can solve for  $\langle m_L^2 \rangle_{\text{st}}$ :

$$\langle m_L^2 \rangle_{\text{st}} = \frac{(H_{\text{PA}}(\tau) + \tau)\mu_2 - \tau(\mu_2 - N\mu^2) \langle m \rangle_{\text{st}}^2}{N\mu^2\tau + H_{\text{PA}}(\tau)\mu_2}. \quad (\text{S256})$$

Finally, we substitute this back into Eq. (S245) and obtain

$$\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{SPA}}(\tau) \left( 1 - \langle m \rangle_{\text{st}}^2 \right), \quad (\text{S257})$$

where

$$H_{\text{SPA}}(\tau) = \frac{\tau H_{\text{PA}}(\tau)(N\mu^2 - \mu_2)}{N\mu^2\tau + H_{\text{PA}}(\tau)\mu_2}, \quad (\text{S258})$$

(the subscript ‘SPA’ of course stands for ‘stochastic pair approximation’). Expanding in  $1/N$  Eq. (S258) can be written as

$$H_{\text{SPA}}(\tau) = H_{\text{PA}}(\tau) \left[ 1 - \frac{\mu_2}{N\mu^2} \frac{\tau + H_{\text{PA}}(\tau)}{\tau} \right] + \mathcal{O}(N^{-2}), \quad (\text{S259})$$

making it transparent that the result from the stochastic pair approximation is a finite-size correction to the expression obtained from the conventional pair approximation.

Due to the terms involving  $H_{\text{PA}}(\tau)$  this solution applies in the same regions as that of the conventional pair approximation, see Sec. S3 C for a detailed discussion of this, as well as Secs. S8 and S10 B for discussions on the  $\tau$  limits in general.

Focusing on positive  $\tau$ , we substitute the solution for  $H_{\text{PA}}(\tau)$  from Eq. (S45) into Eq. (S258) and then expand in powers of  $\tau$ . To lowest order we find

$$H_{\text{SPA}}(\tau) = \left( \frac{N\mu^2}{\mu_2} - 1 \right) \tau + \mathcal{O}(\tau^2). \quad (\text{S260})$$

Thus it is clear that in the limit  $\tau \rightarrow 0$ ,  $H_{\text{SPA}}(\tau) \rightarrow 0$ . This matches what we see in simulations on finite size networks.

In the limit  $\tau \rightarrow \infty$ , Eq. (S259) becomes

$$\lim_{\tau \rightarrow \infty} H_{\text{SPA}}(\tau) = 1 - \frac{\mu_2}{N\mu^2}. \quad (\text{S261})$$

This is of similar form to Eq. (S106) when we studied the annealed approximation. We know the large  $\tau$  regime is dominated by random state changes so there are no correlations between neighbours, hence we expect  $H(\tau) \rightarrow 1$ , as this corresponds to random scattering. Eq. (S261) reproduces this, up to corrections in  $1/N$ .

In Fig. S10 we demonstrate how Eq. (S258) compares to simulation for different size degree-regular (left), Erdős–Rényi (middle), and Barabási–Albert (right) networks. We find convincing agreement across all ranges of  $\tau$ . We note that the approximation becomes inaccurate for very small networks [see Fig. 3(b) in the main paper].

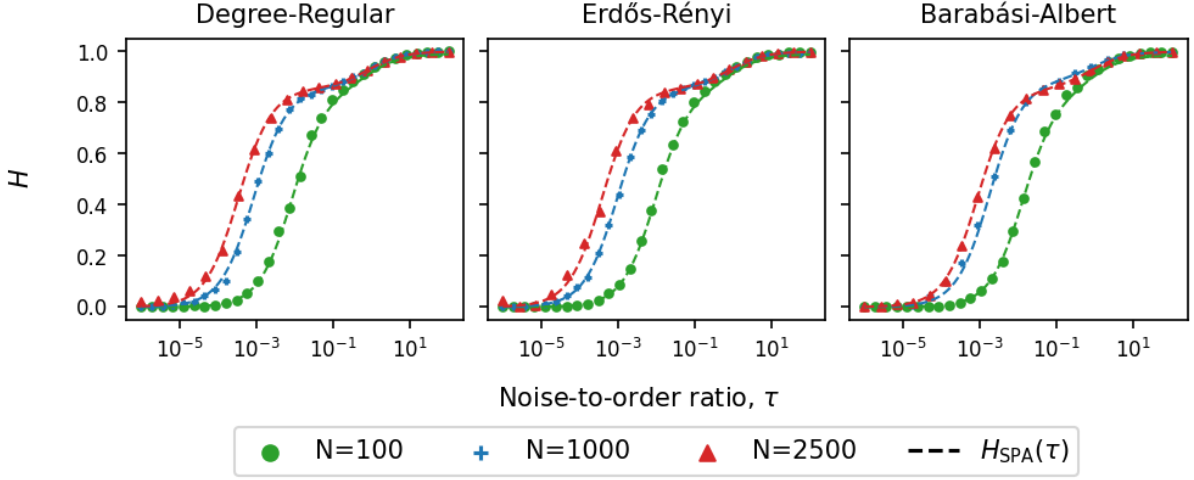


FIG. S10. Plots of  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)]$ , which characterises the scattering, as a function of the noise-to-order ratio,  $\tau$ , for networks of size  $N$ . Degree-regular with  $\mu = 8$  (left), Erdős-Rényi with  $p = 8/N$  (middle), Barabási-Albert with 4 new links established upon adding a new node (right). For all three types of network we have  $\mu \approx 8$  (deviations are due to the finite size of the networks). The dashed lines are the results from the SPA approximation, Eq. (S258). Each marker is the result from averaging 1,000 independent simulations in the steady-state at a particular value of  $\tau$ . Model parameters for a particular  $\tau$  value can be generated via the algorithm detailed in Sec. S10B.

## S8. Scattering and the noise-to-order ratio

### A. Relation between $\mathbb{H}$ and $H(\tau)$

#### 1. Degree of scatter for individual realisations

For individual configurations of the system, the parameter  $\mathbb{H}$  is defined as [Eq.(2)]

$$\mathbb{H} \equiv \frac{\sigma}{\frac{1}{2}(1 - m^2)}. \quad (\text{S262})$$

$\mathbb{H}$  quantifies the ‘scatter’ of a spin configuration, relative to a random configuration with the same magnetisation (same fraction of up-spins). More specifically, the denominator in Eq. (S262) is the (expected) fraction of links that are active if spins of total magnetisation  $m$  are placed randomly on the network. The quantity  $\mathbb{H}$  is not defined if  $m = \pm 1$ , i.e. all spins are in the same state. From an ensemble of independent simulation runs we can measure the quantity  $\langle \mathbb{H}(t) \rangle$ . In computing this average, we discard any configurations where  $\mathbb{H}$  is not defined. Such configurations are very rare provided  $\tau$  is not too small. The exclusion is always implied when we write  $\langle \mathbb{H} \rangle$ .

In Fig. S11 we show  $\mathbb{H}(t)$  (green line) for a single simulation on a Barabási-Albert network of size  $N = 2,500$ , and  $\langle \mathbb{H}(t) \rangle$  obtained from many independent simulations (blue line). The value of  $\mathbb{H}(t)$  fluctuates around  $\langle \mathbb{H} \rangle_{\text{st}}$ . The amplitude of the fluctuations is about one tenth of the mean in this example. Therefore, the value of  $\mathbb{H}$  measured from a single configuration in the stationary state will be well approximated by  $\langle \mathbb{H} \rangle_{\text{st}}$ .

#### 2. Relation between average density of active interfaces and average magnetisation

We define a second order parameter

$$H(\tau) \equiv \frac{\langle \sigma \rangle_{\text{st}}}{\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)}. \quad (\text{S263})$$

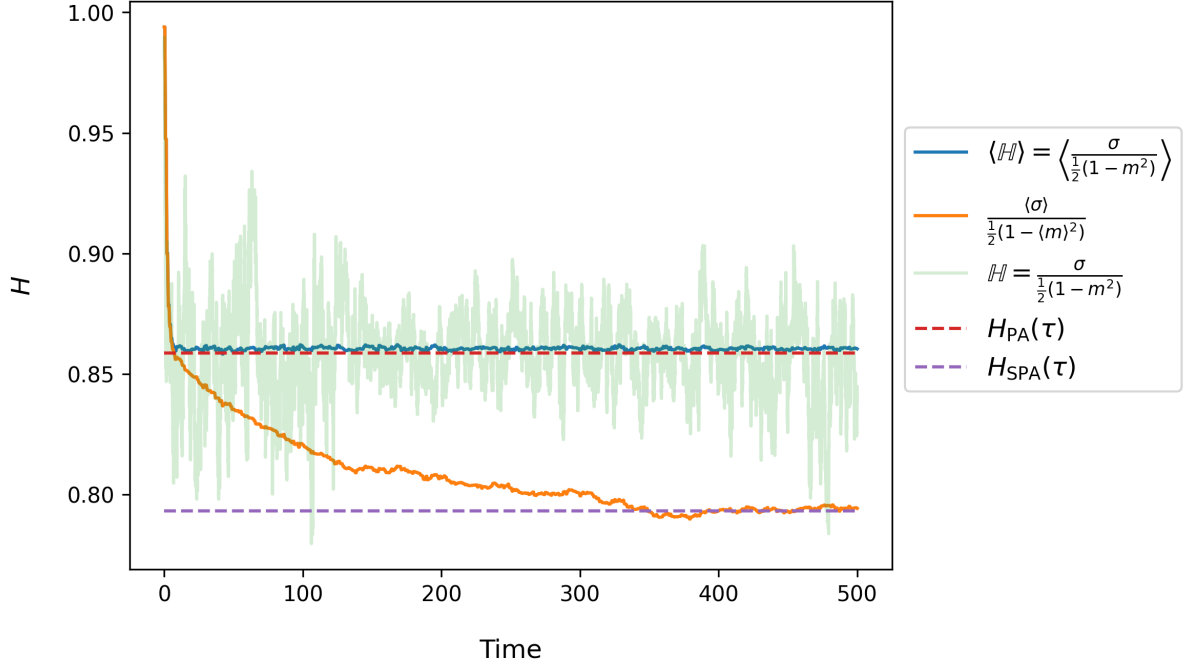


FIG. S11. Time evolution of various order parameters for a Barabási–Albert network with  $\mu \approx 8$  and  $N = 2,500$ . The red and purple dashed lines are theory lines from  $H_{PA}(\tau)$  [Eq. (S45)] and  $H_{SPA}(\tau)$  [Eq. (S258)] respectively. The solid light green line is the measured value of  $\mathbb{H}(t)$  in a single simulation. The solid blue line is  $\langle \mathbb{H}(t) \rangle$  over many such simulations. The solid orange line is  $H(\tau, t)$ , as defined in the text and in the legend, and obtained as an average over many simulations. Data in this example is for  $\tau = 10^{-2}$ .

Even if individual configurations are in ‘consensus’ states ( $m = \pm 1$ ), time averages will lead to  $\langle m \rangle_{st} \neq \pm 1$  (except for pathological choices of the model parameters). We can also define  $H(\tau, t) \equiv \frac{\langle \sigma(t) \rangle}{\frac{1}{2}(1-\langle m(t) \rangle^2)}$ . This quantity is shown as an orange line in Fig. S11 (for  $\tau = 10^{-2}$ ). The data demonstrates that  $H(\tau)$  will in general not take the same value as  $\langle \mathbb{H} \rangle_{st}$  (blue line).

We note in Fig. S11 that, in the stationary state,  $\mathbb{H}$  for a single realisation (green line) fluctuates around  $H_{PA}(\tau)$  while  $H(\tau, t)$  (orange line) approaches  $H_{SPA}(\tau)$ . This is analogous to Fig. 2 in the main paper, where individual trajectories fluctuate around a parabola whose height is set by  $H_{PA}(\tau)$  but where the points defined by  $\langle \sigma \rangle_{st}$  and  $\langle m \rangle_{st}$  sit on a parabola with a different height  $H_{SPA}(\tau)$ .

### 3. $H(\tau)$ can be inferred from measurement $\mathbb{H}$ in individual configurations, for large $\tau$

In Fig. S12 we compare  $\langle \mathbb{H} \rangle_{st} = \left\langle \frac{\sigma}{\frac{1}{2}(1-m^2)} \right\rangle_{st}$  and  $H(\tau) = \frac{\langle \sigma \rangle_{st}}{\frac{1}{2}(1-\langle m \rangle_{st}^2)}$  over a range of values of  $\tau$  for square lattices and Barabási–Albert networks of different sizes. We find that  $H(\tau)$  measured in simulations follows the analytical predictions  $H_{HG}(\tau)$  in the case of finite lattices, and  $H_{SPA}(\tau)$  for the Barabási–Albert network.

We note in Fig. S12 that  $\langle \mathbb{H} \rangle_{st}$  measured in simulations agrees with the analytical predictions for  $H(\tau)$  (and with the measurements of this quantity in simulation) above some  $\tau_C$ . The value of  $\tau_C$  depends on the type and size of the network. Deviations are seen for  $\tau < \tau_C$ . The value of  $\tau_C$  decreases as  $N$  becomes larger, see e.g. the triangle markers in Fig. S12(a).

Analytical functions for the order parameter  $H(\tau)$  can be derived for many networks. These are the functions  $H_X(\tau)$  [Eqs. (S45), (S105), (S183) and (S258)]. These functions depend only on properties of the network (e.g. the average degree  $\mu$ ), and their argument is  $\tau$  [see Sec. S8B]. This, together with the above, provides a method to infer  $\tau$  from snapshots of spin configurations in the steady-state.

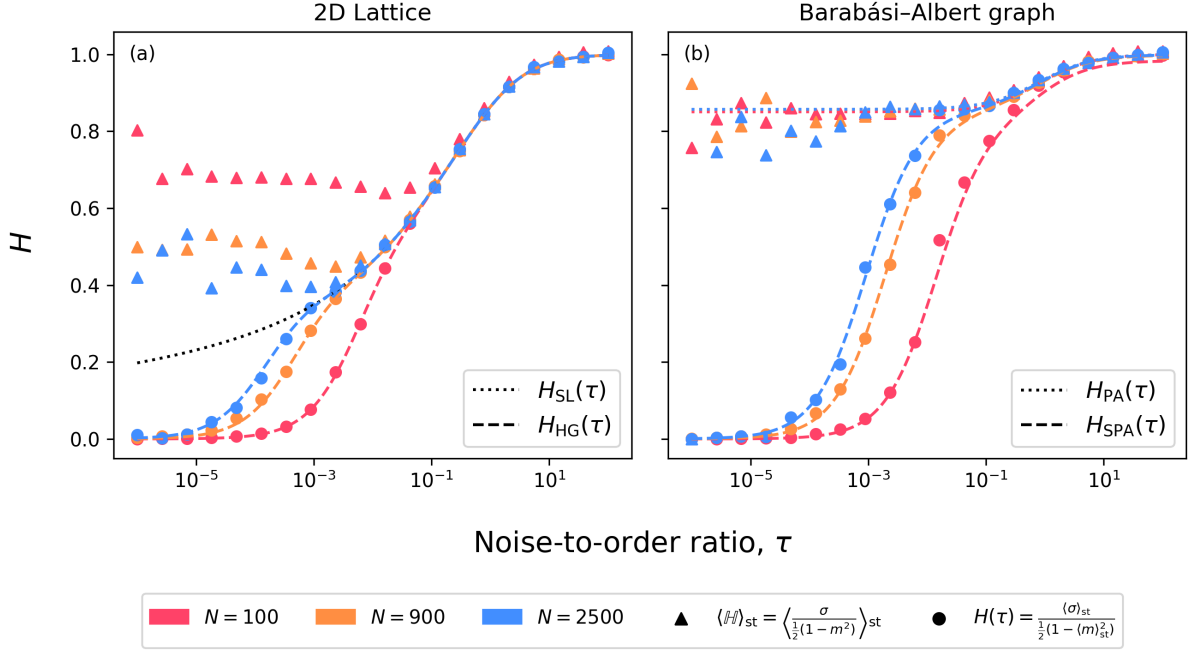


FIG. S12. Plots comparing  $\langle \mathbb{H} \rangle_{\text{st}}$  (triangles) and  $H(\tau)$  (circles), from Eqs. (S262) and (S263), in simulations on square lattices with periodic boundary conditions (left) and Barabási–Albert networks with  $\mu \approx 8$  (right), of varying size  $N$ . Left panel: dotted line is the  $H_{\text{SL}}(\tau)$  theory line for an infinite square lattice [Eq. (S120)], dashed lines are the  $H_{\text{HG}}(\tau)$  theory lines for homogeneous networks [Eq. (S183)]. Right panel: dotted lines are  $H_{\text{PA}}(\tau)$  theory lines for infinite uncorrelated networks [Eq. (S45)], dashed lines are the  $H_{\text{SPA}}(\tau)$  theory lines for large but finite uncorrelated networks [Eq. (S258)]. In both panels we find that there is much more stochasticity in  $\langle \mathbb{H} \rangle_{\text{st}}$  (triangles) than in  $H(\tau)$  (circles). This is because if the system reaches consensus (all spins up or down),  $\mathbb{H}_{\text{st}}$  cannot be measured and the sample is discarded. This happens the majority of the time at low  $\tau$  values meaning many more simulations are needed to measure  $\langle \mathbb{H} \rangle_{\text{st}}$  than  $H(\tau)$ .

#### 4. Range of possible $H(\tau)$ values

Since  $m$  is bound on  $[-1, 1]$ , and  $\sigma$  on  $[0, 1]$ ,  $H(\tau)$  can, in principle, take any real value greater than or equal to 0. However, not all combinations of  $\langle m \rangle_{\text{st}}$  and  $\langle \sigma \rangle_{\text{st}}$  are realised by the model dynamics. In practice the maximum value of  $H(\tau)$  that can be achieved depends on the topology of the network. For general networks, finding the assignment of states  $\pm 1$  for each node leading to the largest possible number of active links is closely related to the maximum-cut problem, which is known to NP-hard [S25].

### B. Noise-to-order ratio, $\tau$

#### 1. Definition and interpretation

The noise-to-order ratio,  $\tau$ , is given by [see Eq. (S46)]

$$\tau = 2 \frac{h'_+ + h'_- + v'_+ + v'_-}{(h_+ + h_-) - (h'_+ + h'_-)}. \quad (\text{S264})$$

#### Interpretation of the numerator

To interpret the numerator we imagine a fully ordered spin configuration, that is, all spins up or all spins down. For example, if all spins are up, then the only possible events are spontaneous down-flips or down-flips due to incorrect copying. As seen from Eqs. (S2) and (S3) the total event rate for the former is  $v'_- N$  and for the latter  $h'_- N$ . So the total event rate, starting from the all-up-state is  $(v'_- + h'_-) N$ . Similarly, the total

event rate starting from the all down configuration is  $(v'_+ + h'_+)N$ . Up to the factor of  $N$ , we can thus think of the numerator in Eq. (S264) as the average event rate in a fully ordered state. Any event in such a state moves the system towards disorder.

*Interpretation of the denominator*

We now show that the denominator describes the initial rate of the build-up of nearest-neighbour correlations if the system is started from a fully uncorrelated state with  $m = 0$ . In other words, we assume that, in the initial state, each spin is up or down with equal probability, and independently of its neighbours. Then we calculate  $\frac{d}{dt} [\langle ss' \rangle - \langle s \rangle \langle s' \rangle]$ .

If a give node is the state  $+1$  and if  $n$  of its  $k$  nearest neighbours are in state  $-1$ , then the rate at which it flips to  $-1$  is

$$T_{n,k,t}^- = h_- \frac{n}{k} + h'_- \frac{k-n}{k} + v'_-. \quad (\text{S265})$$

Spontaneous down flips in one node (described by  $v'_-$ ) cannot affect correlations between neighbours. In fact, in the fully random state with  $m = 0$  such a spontaneous flip will, on average, convert as many active interfaces into inactive interfaces as the other way round. Therefore, we do not consider this term any further for the purposes of calculating the build-up of correlations.

The first two terms in Eq. (S265) describe the following: A random neighbour of the node is chosen. If that neighbour is in the down-state (this happens with probability  $1/2$  in the fully uncorrelated state) the focal node flips with rate  $h_-$ . The pair of focal node and neighbour change from  $+-$  to  $--$ . The change in  $ss'$  is thus  $+2$ . If instead the neighbour is in the up-state (which also happens with probability  $1/2$ ), the focal node flips with rate  $h'_-$ . The pair of focal node and neighbour change from  $++$  to  $-+$ , and the change in  $ss'$  is  $-2$ .

We note that interfaces of the focal node with any other nearest neighbours (other than the designated interaction partner) are equally likely to change from active to inactive or from inactive to active, given that we are starting from the fully uncorrelated state. These interfaces therefore do not contribute to any build-up of correlations.

The total expected rate of change of  $\langle ss' \rangle$  due to events described by Eq. (S265) is thus  $\frac{1}{2}h_- \cdot (+2) + \frac{1}{2}h'_- \cdot (-2) = h_- - h'_-$ .

Similarly, if the focal node is in state  $-1$  and  $n$  is the number of neighbours in the up-state, then the rate at which it flips to state  $+1$  is

$$T_{n,k,t}^+ = h_+ \frac{n}{k} + h'_+ \frac{k-n}{k} + v'_+. \quad (\text{S266})$$

We can make an analogous argument as above to show that the total expected rate of change of  $\langle ss' \rangle$  due to events described by Eq. (S266) is thus  $h_+ - h'_+$ .

Finally, we note that

$$\frac{d}{dt} (\langle s \rangle \langle s' \rangle) = \langle s \rangle \frac{d}{dt} \langle s' \rangle + \langle s' \rangle \frac{d}{dt} \langle s \rangle = 0 \quad (\text{S267})$$

in the initial state in which  $m = 0$ , i.e.  $\langle s \rangle = 0$ .

Putting everything together we therefore find that

$$\frac{d}{dt} \langle ss' \rangle \propto h_- - h'_- + h_+ - h'_+. \quad (\text{S268})$$

Thus, we have shown that the denominator of the expression in Eq. (S264) describes (up to constant pre-factors) the initial rate of the build-up of nearest neighbour correlations if the dynamics is started from a fully disordered state in which each spin is up or down with probability  $1/2$ , and independently of its neighbours.

2. Range of possible values for  $\tau$

The definition in Eq. (S264) is of the form

$$\tau = 2 \frac{a+b}{c-a}, \quad (\text{S269})$$

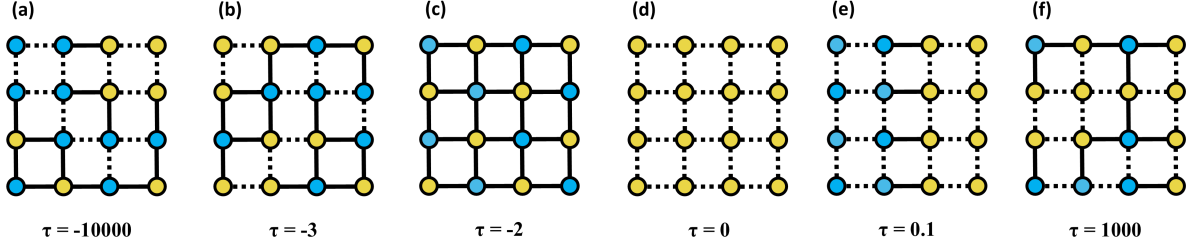


FIG. S13. Snapshots of the spin distribution of a single realisation on a grid network after running the model dynamics for a long time, for different values of  $\tau$ . Nodes in state  $+1$  are blue. Nodes in state  $-1$  are yellow. Links connecting different state nodes are active (solid). Links connecting same state nodes are inactive (dashed).

with  $a, b, c, d \geq 0$ .

It is clear from this definition that all values  $\tau \geq 0$  can be realised through a suitable choice of  $a, b, c, d$ .

The quantity  $\tau$  takes negative values, if  $a > c$ . In this case we have

$$-\tau^{-1} = \frac{1}{2} \frac{a - c}{a + b}, \quad (\text{S270})$$

and  $0 < a - c \leq a \leq a + b$ . Thus  $-\tau^{-1} \leq 1/2$ , and hence  $\tau \leq -2$ .

We conclude that  $\tau$  can only take values in the ranges  $\tau \geq 0$  or  $\tau \leq -2$ .

### C. Behaviour in different regimes

In Fig. S13 we show snapshots of the spin configuration in the stationary state, obtained from simulations of the model on a  $4 \times 4$  square grid.

#### 1. The limit $\tau \rightarrow -\infty$

The limit  $\tau \rightarrow -\infty$  emerges in two different ways:

- (i) One or both of  $v'_\pm$  tend to infinity, while  $h'_+ + h'_- > h_+ + h_-$ ; or
- (ii) The denominator of Eq. (S264) approaches zero from below.

Typical configurations of the system will be mostly random, but there will be some small anti-correlation between nodes so the number of active links is slightly greater than random. We find  $H(\tau) \rightarrow 1^-$  in this case. This case is illustrated by the configuration in Fig. S13(a), for which we find  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m^2)] = 1.02$ .

#### 2. Finite negative values of $\tau < -2$

The copying process is mostly unfaithful, but there can also be faithful copying and/or spontaneous spin flips. There will be some anti-correlation between neighbouring spins, and thus, for a given trait frequency  $m$ , one expects more active interfaces than a configuration in which these spins are placed randomly on the network. Hence, we expect values of  $H(\tau) > 1$ .

A case in which  $\tau$  is negative and finite is illustrated by the configuration in Fig. S13(b), for which  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m^2)] = 1.27$ .

3. *The limit  $\tau \rightarrow -2^-$*

From the definition of  $\tau$  in Eq. (S264) we have

$$\tau = -2 \frac{h'_+ + h'_- + v'_+ + v'_-}{(h'_+ + h'_-) - (h_+ + h_-)}. \quad (\text{S271})$$

Hence  $\tau = -2$  when either  $h'_+ + h'_-$  diverges and the  $v'_\pm, h_\pm$  remain finite, or if the  $v'_\pm$  and  $h_\pm$  all vanish. In either case, the dynamics is dominated by unfaithful copying from neighbours, i.e. the model is type of ‘contrarian voter model’, in which a spin copies the state opposite to that of a neighbour. This creates strong anti-correlations between the states of neighbouring spins, and  $H(\tau)$  will take its maximum value which is specific to the network.

This scenario is illustrated in Fig. S13(c), where we show a perfect ‘checkerboard’ pattern seen in simulations on a small square grid. For this pattern we have  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m^2)] = 2$ .

4. *The limit  $\tau \rightarrow 0^+$*

The limit  $\tau \rightarrow 0^+$  can occur in one of two ways:

- (i) The numerator in Eq. (S264) goes to zero (from above) and, at the same time,  $h_+ + h_- > h'_+ + h'_-$ , so that denominator is positive; or
- (ii) One has  $h_+ + h_- \rightarrow \infty$ .

In the former case we have  $h'_\pm \rightarrow 0$  and  $v'_\pm \rightarrow 0$ . There is no transmission noise. In the latter case faithful copying from neighbouring spins dominates the dynamics.

In either case the number of active links is low and we expect configurations in which either all spins are up or all spins are down, as illustrated in Fig. S13(d). Due to the convergence to a fully ordered state, the ratio  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m^2)]$  cannot be defined for single realisations. However all realisations have  $\sigma = 0$  eventually, and we measure  $H(\tau) = \langle \sigma \rangle_{\text{st}} / [\frac{1}{2}(1 - \langle m \rangle_{\text{st}}^2)] = 0$ .

5. *Finite  $\tau > 0$*

For  $\tau > 0$  and finite, there are some unfaithful events occurring, and the rate of faithful horizontal events is greater than the total rate of unfaithful horizontal events, i.e.  $h_+ + h_- > h'_+ + h'_-$ . This leads to steady-state situations where  $0 < H(\tau) < 1$ . The case is illustrated in Fig. S13(e), where we measure  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m)^2] = 0.33$ .

6. *The limit  $\tau \rightarrow \infty$*

This limit emerges in two different ways:

- (i) The denominator in Eq. (S264) remains positive, and either  $v'_+$  or  $v'_-$  (or both) tend to infinity; or
- (ii) The  $v'_\pm$  remain finite, but the denominator in Eq. (S264) diverges, i.e.,  $h_+ + h_- = h'_+ + h'_-$ .

In the former case, the system is dominated by random spin flips. In the second scenario, correct and incorrect copying from neighbouring nodes occur at the same rates. This means that the copying spin takes a state that is effectively unrelated to that of the state it is trying to copy.

In neither of these two situations correlations between neighbouring states can build up, and so the resulting configurations are effectively random. In general  $H(\tau) \rightarrow 1^-$ . This case is illustrated in Fig. S13(f), where we find  $\mathbb{H} = \sigma / [\frac{1}{2}(1 - m^2)] = 0.98$ .



## 7. Illustration and test against simulations

Fig. S14 shows  $H_{\text{SL}}(\tau)$ , from Eq. (S120), as an example, for both positive and negative values of  $\tau$ .

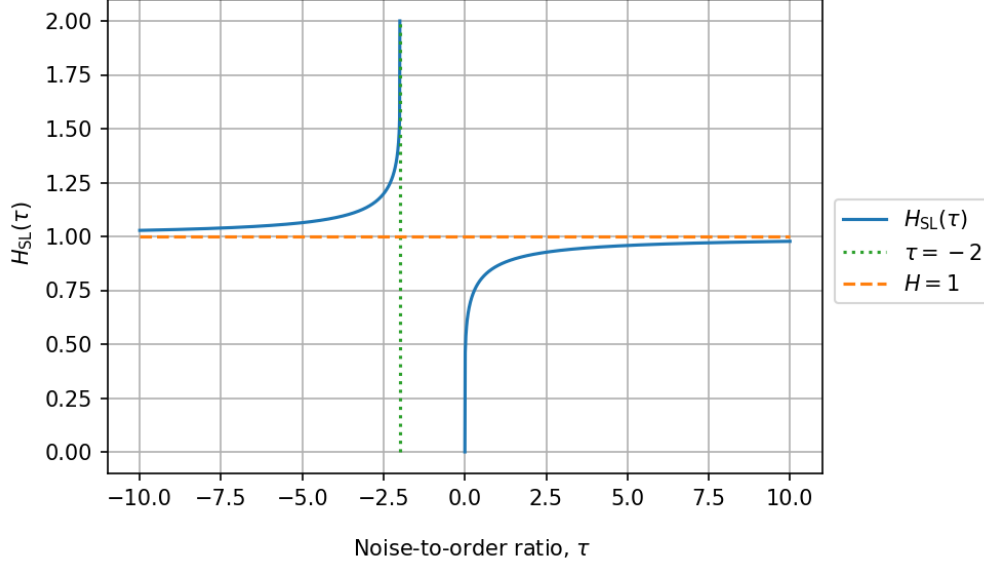


FIG. S14. Plot of  $H_{\text{SL}}(\tau)$  from Eq. (S120) over a range of positive and negative values of  $\tau$ . The horizontal orange dashed line shows  $H = 1$ , and the vertical green dotted line is at  $\tau = -2$  for reference.

In Fig. S15 we show simulation results to illustrate the effects of negative values of  $\tau$ , and to demonstrate the validity of our analytical description. Specifically, simulations are on a square lattice and we show plots of  $\langle \sigma \rangle_{\text{st}}$  as a function of  $\langle m \rangle_{\text{st}}$ . As before these curves are obtained by varying the detailed model parameters, while keeping  $\tau$  fixed [see Sec. S10 B]. We also show the analytical predictions for the square lattice, i.e.  $\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{SL}}(\tau) (1 - \langle m \rangle_{\text{st}}^2)$  with  $H_{\text{SL}}(\tau)$  from Eq. (S120).

For  $\tau = 1$ , the model dynamics contains some transmission noise, but copying from neighbours is more likely to be faithful than unfaithful. We have  $H(\tau) < 1$ .

For very large positive  $\tau$ , noise dominates, and the resulting configurations at a given  $m$  are effectively random, so  $H(\tau) = 1$ .

For  $\tau = -3$ , copying events are mostly unfaithful (that is, spins take the state opposite to the one they are ‘copying’). There is also spontaneous flipping, and some faithful copying. Overall, some degree of anti-correlation between neighbours ensues, and so  $H(\tau) > 1$ .

The stationary average density of active interfaces at the apexes of the three parabolas (located at  $\langle m \rangle_{\text{st}} = 0$ ) are given by  $\langle \sigma \rangle_{\text{st}} = \frac{H(\tau)}{2}$ , and are smaller than 0.5 for  $\tau = 1$ , equal to 0.5 for  $\tau \rightarrow \infty$ , and larger than 0.5 for  $\tau = -3$ .

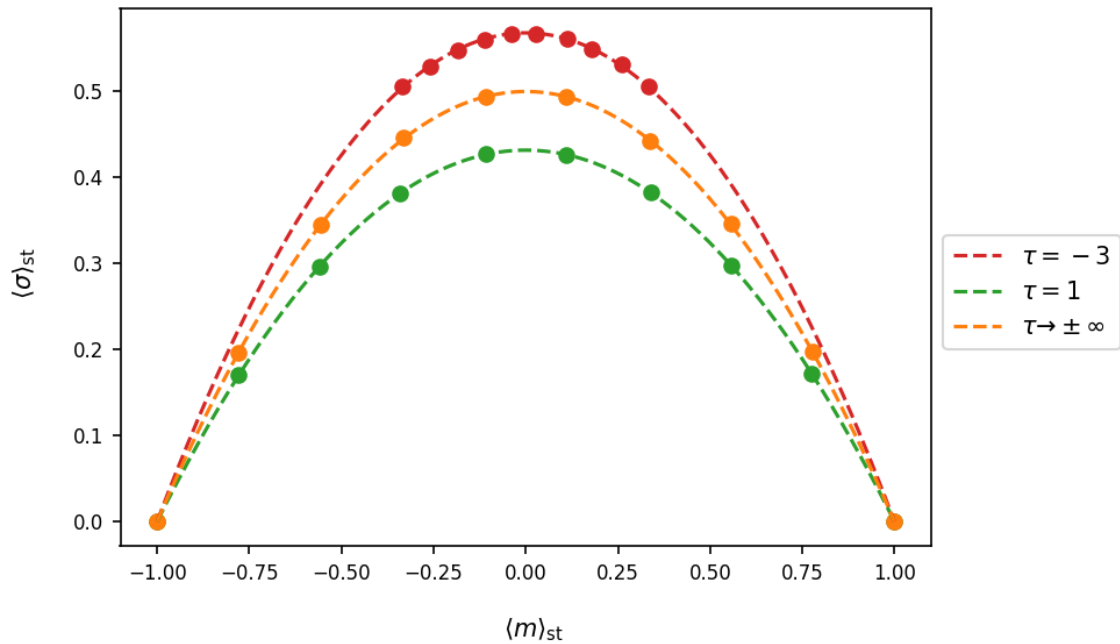


FIG. S15. Parabolas  $\langle \sigma \rangle_{\text{st}} = \frac{1}{2} H_{\text{SL}}(\tau)(1 - \langle m \rangle_{\text{st}}^2)$  for the model on an infinite square lattice, and for different values of  $\tau$ . Dashed lines are from the theory [Eq. (S120)], markers are the results from averaging the steady-state values of 10 independent simulations on a lattice with  $N = 10,000$  nodes (periodic boundary conditions). For  $\tau = -3$  the algorithm we use to generate model parameters [see Sec. S10 B] limits the values of the stationary magnetisation that we can probe. More specifically, for parameters generated by this algorithm,  $\langle m \rangle_{\text{st}}$  can only take values in the range given in Eq. (S284). For the other cases  $\langle m \rangle_{\text{st}}$  can take any value in the interval  $[-1, 1]$ . The dashed line labelled  $\tau \rightarrow \pm \infty$  corresponds to  $H(\tau) = 1$ , the orange markers are from simulations with  $\tau = 1,000$ .

## S9. Analysis of data from the World Atlas of Language Structures

### A. Further details on the speaker-level processes underpinning the model

When our general setup is used to model language change, each node in the network is a language and the dynamics happens at the level of languages as well. Thus, the model is a coarse-grained description of the more microscopic speaker-level processes. In this section we briefly describe what these processes are.

At the microscopic level, language change is driven by the behaviour of individual speakers in the processes of language acquisition and language use. Endogenous change (captured by spontaneous spin-flip in our model, i.e., the ‘vertical’ events) arises when a speaker or group of speakers fail to replicate a trait of the ambient language, for example children in the process of acquiring language, or adults in the process of using it. This introduces a new trait that then propagates throughout the speech community. At this level, traits are best thought of as relative rates of use of competing options, and changes between traits typically follow sigmoid patterns of transition between steady states in which opposite options are predominant. The propagation of individual innovations is then modulated by processes of social selection that are sensitive to the social embedding of the new trait (i.e. what demographic groups use it). For a recent approach, with references to previous literature, see e.g. [S26].

Change by contact occurs when languages interact. At one end of the spectrum, the interaction may involve fully balanced bilingual speakers who are equally competent in both languages; at the other, it may involve native speakers of one language who have limited knowledge of another language as a foreign language; in the middle, there are different mixtures of these and other intermediate possibilities. Horizontal transfer occurs when a trait of language A influences a speaker’s use of language B. In the simplest case, a trait of A is introduced in the speaker’s use of B, but the change in B may not be a straightforward copy of a trait of A; it may, for example, be a simplification. This corresponds to the two horizontal possibilities in our model: faithful horizontal transmission and unfaithful horizontal transmission.

In general, other things being equal, spatial proximity increases the chance that two languages are engaged in a mutual horizontal interaction, since the effect of a language A on another language B grows with the proportion of speakers of B who know A, with their competence in A, and with the frequency with which the use A. For structural linguistic traits of the sort involved in our study, contact effects arise only under fairly intense interaction, of the type that occurs in bilingual communities or in border areas. For a survey of questions pertaining to horizontal interactions between languages, see e.g. [S27].

## B. Dataset

In this section we provide further information on the analysis of data from WALS (The World Atlas of Language Structures) [S28] in the main paper. WALS contains information on 2,662 human languages, most importantly the co-ordinates of where that language is most densely spoken. The details on how to determine the exact co-ordinations of a language are complex, see Sec. 3.4 in [S29]. We will refer to traits by their WALS-ID, which is a combination of numbers and letters such as ‘11A’, ‘143F’ etc.

Languages possess a number of language traits (also known as ‘typological features’). WALS details 192 such traits. Not all traits are recorded for all languages in WALS. We focus on a curated set of 35 language traits which can be quantified in a sufficient set of languages, and which are ‘binarisable’, that is, if the trait can be quantified in the language it can meaningfully be classified as present or absent. The set of traits and languages is that of [S1], where further details can be found. Ultimately, we consider 2,248 languages from WALS, these are those for which at least one trait is quantifiable in binary form. This means that not all nodes (languages) have a well defined binary state for all 35 traits.

Binary traits can all be naturally mapped to our theoretical set-up. Nodes represent languages, the spin states indicate whether that language possesses or lacks a specific language trait, links between nodes indicate interaction.

The relevant data to perform this analysis can be found at the following GitHub repository [S24]. This includes details about traits, distances between languages, and information on which languages can be assigned a binary state for which traits.

## C. Empirical noise-to-order ratio

To determine the density of active interfaces for each of the 35 traits from the WALS data we need to decide how the languages are connected. We do not know the true interaction network. We therefore make the following choice: two languages  $A$  and  $B$  are connected if  $B$  is among the 10 geographically closest neighbours of  $A$ , or if  $A$  is among the 10 geographically closest neighbours of  $B$ . This creates an undirected network.

With this method, languages can end up having more than 10 neighbours in the network. For example, start by considering  $A$ . We want to determine which other languages it connects to. We find the 10 geographically nearest languages of  $A$  and connect it to those languages. Assume that  $B$  is one of these languages. We then consider  $B$ . It already has one connection to  $A$ . We then find its 10 geographically nearest languages. It is possible that  $A$  is not within this set. We connect  $B$  to these 10 languages.  $B$  now has 11 neighbours in the network.

The choice of 10 nearest neighbours is somewhat arbitrary, however, the results of [S1] for example have been shown to be robust against variation in neighbourhood size.

The degree distribution for the above network is shown in Fig. S16. The network is not regular, neither is it uncorrelated as it is likely that two neighbours of a given node are connected to one another. However the degree distribution is fairly tight, so we can still apply the various analytical results to good effect.

To determine an empirical noise-to-order ratio for a particular binarisable trait, we assign states  $\pm 1$  to those nodes (languages), for which the binary state of the trait is defined. The state  $\pm 1$  indicates whether the language represented by the node posses or lacks the trait. If the trait is not quantifiable on a given node, then no spin state is assigned to the node. Taking into account only nodes with states  $\pm 1$ , we then determine the trait frequency  $\rho$ , and, using the putative interaction network, the density of active links,  $\sigma$ . From this we obtain  $\mathbb{H} = \sigma/[2\rho(1 - \rho)]$  as a quantifier of the spatial scattering. Finally, the different analytical functions,

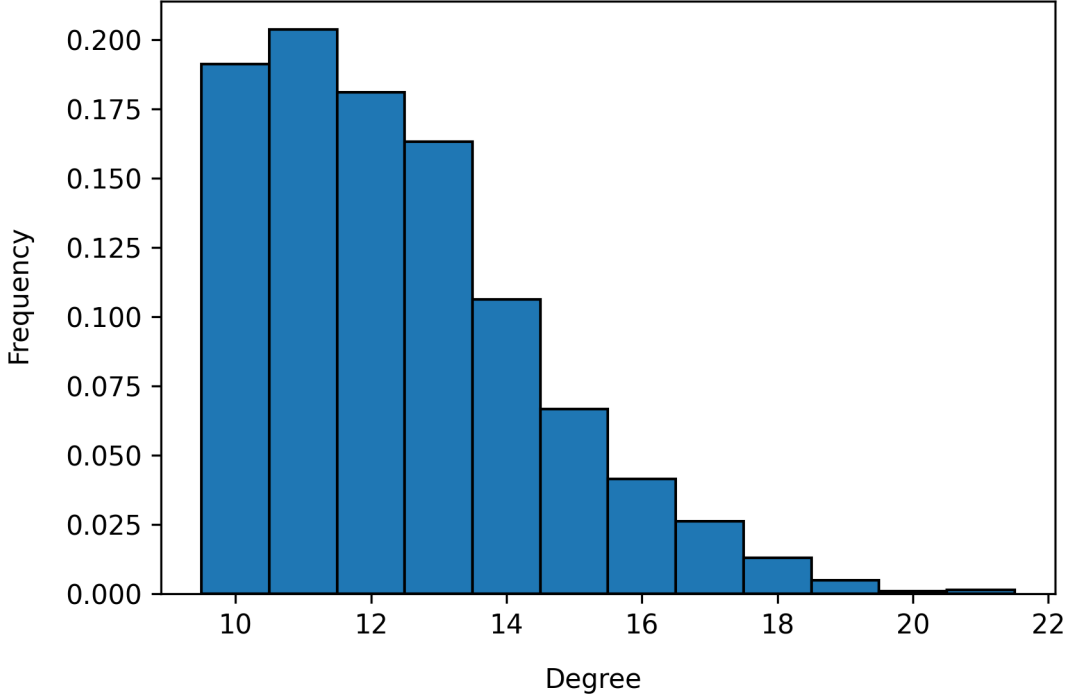


FIG. S16. Degree distribution of our approximation to the real-world network of language interactions (see text).

$H_{AA}(\tau)$  [Eq. (S105)],  $H_{SL}(\tau)$  [Eq. (S120)],  $H_{HG}(\tau)$  [Eq. (S183)] and  $H_{SPA}(\tau)$  [Eq. (S258)], can be used to infer values for the noise-to-order ratio,  $\tau_X$ . The  $X$  subscript stands for one of AA, SL, HG, or SPA depending on the specific  $H_X(\tau)$  function that was used.

Using this procedure it is sometimes possible to obtain  $\mathbb{H} = \sigma/[2\rho(1 - \rho)] > 1$ . As discussed in Sec. S8 this would lead to negative values of  $\tau$  values. Such occurrences are rare in our dataset (less than 1%), so we discard these cases.

#### D. Comparison against longitudinal stability estimate

To validate our results we compare to a phylogenetic method which does not take into account any spatial data. This was done in [S30] using a principal component analysis, and in particular the first component PC1. Higher PC1 values corresponding to greater instability. PC1 values were only measured for a subset of the 35 traits that we consider.

In Fig. S17 (left) we show regression plots of PC1 against  $\tau_X$ . The different rows correspond to different  $X$  and thus use different analytical forms  $H_X(\tau)$ , as described above and indicated in the figure.

A similar but restricted analysis was carried out in [S1]. This was solely based on  $H_{SL}(\tau)$  from the theory for infinite square lattices. That earlier analysis identified a small set of outlier traits for which the estimates of  $\tau$  and Dediu's PC1 gave conflicting indications regarding relative stability [S30] (see also [S31]). Some of these outliers could be justified based on known properties of some of the WALs traits, see the discussion in [S1] for details.

We have carried out a similar identification of outliers to test robustness against the use of different analytical forms  $H_X(\tau)$ . Outliers are determined by recursively pruning the dataset and removing the trait that provide the largest increase in Pearson correlation coefficient  $r$  (measuring linear correlation [S32]). Results are shown in the right-hand panels of Fig. S17. Independent of the analytical form used for the function  $H_X(\tau)$  we find the same set of outliers, namely WALs traits 8A, 11A, 107A and 57A. This is the same set as found in [S1], confirming the robustness against variation of the assumed network structure [reflected by the choice

of analytical form of  $H_X(\tau)$ ].

After outlier removal we find strong correlation between the inferred values of  $\tau_X$  and Dediu's PC1, with an  $r$  consistently above 0.8, and  $p$ -value  $< 0.01$ , for all theories  $X$ .

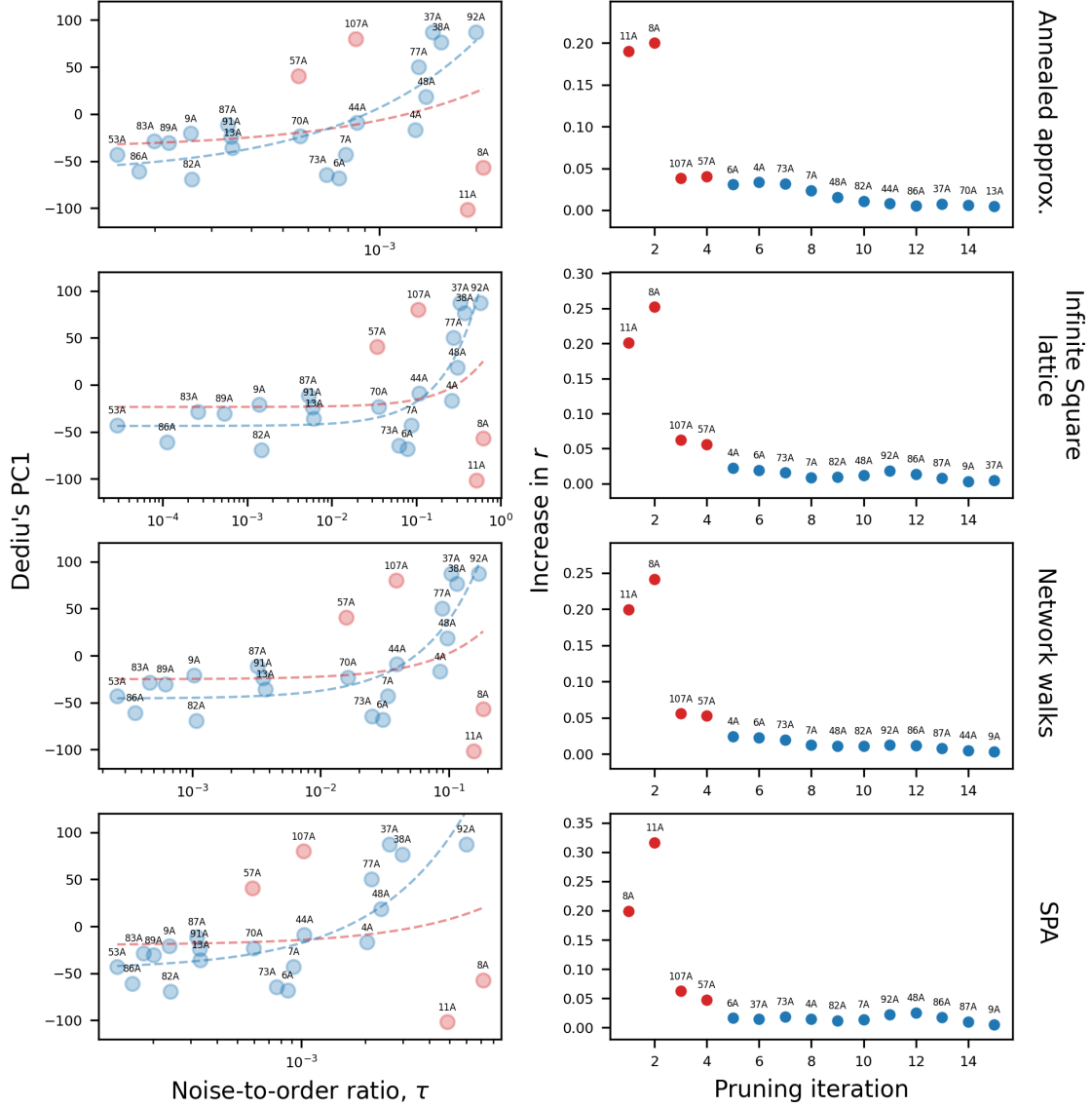


FIG. S17. Panels on the left show regression plots of Dediu's PC1 [S30] against the noise-to-order ratio,  $\tau$ . Red points are deemed outliers (see text). The red and blue dashed lines are the best linear fit before and after outliers are removed (we draw attention to the logarithmic scale on the horizontal axis, making the fits look curved). The different rows correspond to different choices of the analytical form of  $H(\tau)$ , as indicated on the right-hand side. From top to bottom we have used the annealed approximation [Sec. S4], the infinite 2D square lattice [Sec. S5], the network walk method [Sec. S6] and the stochastic pair approximation [Sec. S7] After outlier removal the  $r$  values are 0.813, 0.857, 0.848, 0.821 from top to bottom, all with  $p$ -value  $< 0.01$ . The right column illustrates the outlier removal process. At each pruning iteration the point that gives the largest increase in  $r$  is removed. For clarity, we only show the first 15 iterations.

## E. Bootstrapping and rankograms

### 1. Construction of rankograms

The method in Sec. S9 C can be used to rank the 35 traits in terms of noise-to-order ratio, which in turn can be regarded as a proxy for stability.

Our analysis is based on the WALS dataset, involving  $N = 2,248$  languages, each with at least one binarisable trait among our selected set of 35 traits. In reality, there are other languages for which we do not have the binarisable data. This can be due to lack of observation, or WALS simply is missing the data. We can therefore think of the WALS dataset as an effective random sub-sample of all human languages. To quantify the uncertainty associated with this, we use bootstrap re-sampling [S33].

In our context, bootstrapping involves sampling  $N = 2,248$  languages with replacement from the original WALS dataset. This gives rise to an (assumed) interaction network as described in Sec. S9 C. This network is specific to the bootstrap sample. We note that it is possible to select the same language multiple times. This leads to multiple identical languages with geographical distance zero, thus they are essentially guaranteed to be connected. This represents potentially very close neighbouring languages, not included in WALS, which have the same trait values.

For each sample, we can perform the procedure in Sec. S9 C to calculate a noise-to-order ratio  $\tau_X$  for each trait (discarding nodes in the network for which the status of a given trait cannot be quantified in binary form). Repeating this many times, we generate distributions of  $\tau_X$  for each trait.

To produce rankograms for a chosen  $X$  we can then draw a value of  $\tau_X$  for each trait from the respective distributions. The traits can then be ranked based off of these values. This is then repeated many times and thus a distribution of ranks is constructed for each trait. These are the rankograms shown in Figs. 4 in the main paper, and in S18 in this supplement. These are histograms where the  $i^{\text{th}}$  bar indicates the probability that the trait represented by that rankogram ranks in the  $i^{\text{th}}$  position amongst all traits.

We observe that the traits on the extreme ends, such as 130A or 53A, have very tight rankograms. This means that we can be confident that those traits will rank in those positions amongst all traits. We are less confident about the traits with broad distributions, such as 100A or 48A.

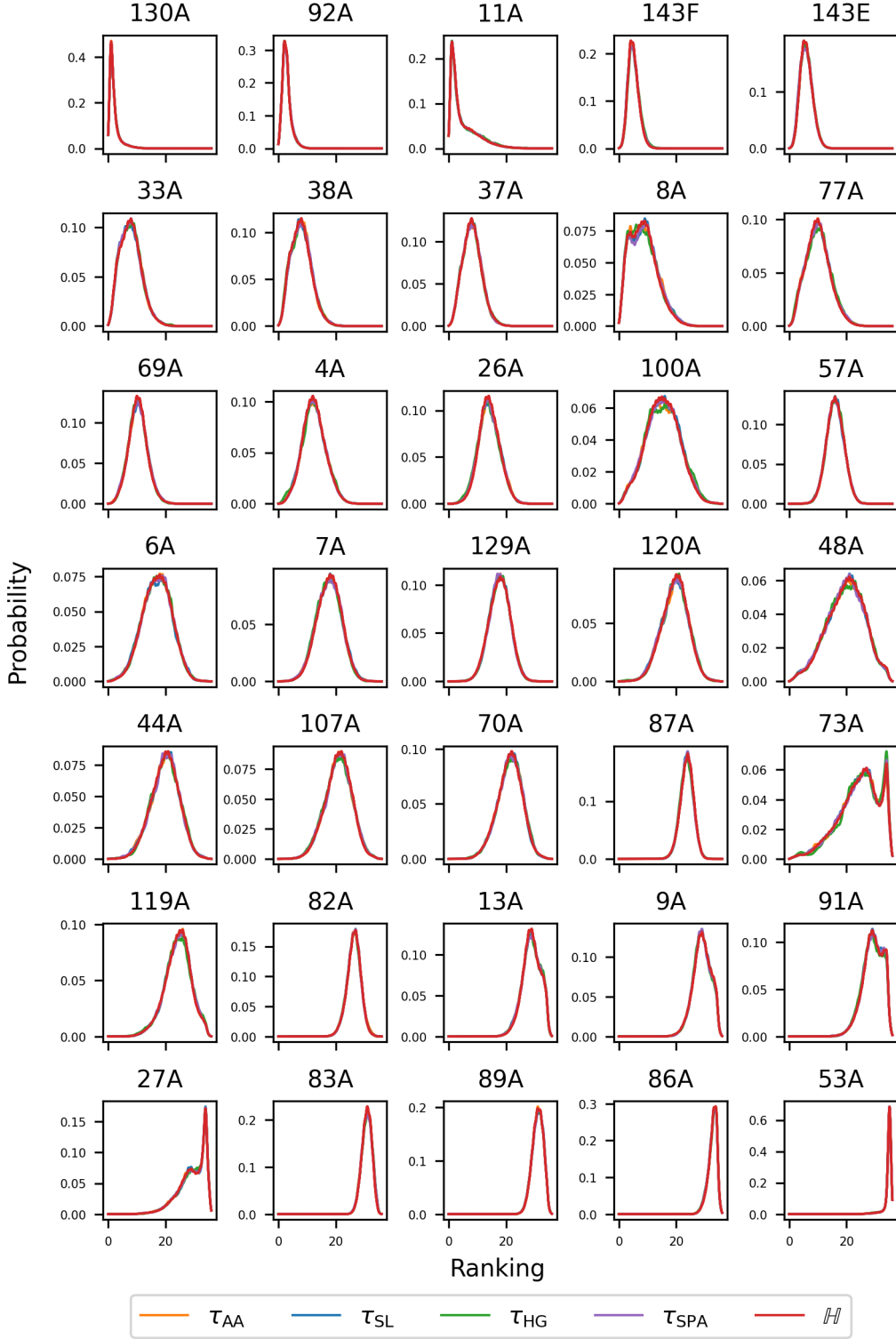


FIG. S18. Rankograms for all 35 language traits for 10,000 bootstrap samples. We only show Gaussian kernel density estimates of the distributions for clarity. Each rankogram has five distributions. Four of which are based on inferring values of  $\tau_X$  from the theories  $H_{AA}(\tau)$  [Eq. (S105)],  $H_{SL}(\tau)$  [Eq. (S120)],  $H_{HG}(\tau)$  [Eq. (S183)] and  $H_{SPA}(\tau)$  [Eq. (S258)]. The fifth is from ranking raw  $\mathbb{H}$  values calculated using Eq. (2). The height at a given rank is the probability a trait is ranked in that position amongst all traits.

2. Distributions of  $\tau_X$ 

From the rankograms in Fig. S18 we see that the rankings produced from  $\mathbb{H}$  and those produced from  $\tau_X$  using each of the theories  $X$  are very similar. This is a non-trivial result and is not a consequence of the fact that all the  $H_X(\tau)$  functions are monotonic increasing functions of  $\tau$  as explained below.

In a single bootstrap sample, we have a set of languages and construct one interaction network. We then measure 35 values of  $\mathbb{H}$ , one for each trait, using the procedure in Sec. S9 C. These  $\mathbb{H}$  values will have a certain ranking. With the single interaction network we construct the functions  $H_X(\tau)$ , one for each theory  $X$ . For a chosen  $X$  the 35  $\mathbb{H}$  values can be mapped to 35  $\tau_X$  values. Since all  $H_X(\tau)$  are monotonic increasing functions of  $\tau$ , ordering the traits based on  $\tau_X$  will be the same as ordering based on  $\mathbb{H}$ , for all theories  $X$ .

However, the procedure for constructing rankograms is different in that we do not perform any rankings until distributions of  $\tau_X$  have been constructed for each trait. To construct distributions of  $\tau_X$  for a trait we repeat the procedure in Sec. S9 C many times. Each bootstrap sample consists of a different set of languages, and thus a different interaction network. Focusing on two traits  $A$  and  $B$ , we measure the corresponding values  $\mathbb{H}^A = \sigma^A/[2\rho^A(1-\rho^A)]$  and  $\mathbb{H}^B = \sigma^B/[2\rho^B(1-\rho^B)]$  in each sample. This gives distributions for  $\mathbb{H}^A$  and  $\mathbb{H}^B$ . For each bootstrap sample we also map the empirical values of  $\mathbb{H}^A$  and  $\mathbb{H}^B$  to values  $\tau_X^A$  and  $\tau_X^B$ , using one of the analytical functions  $H_X(\tau)$ . This gives us distributions of  $\tau_X^A$  and  $\tau_X^B$ .

Since the interaction network is different in each bootstrap sample, the function  $H_X(\tau)$  will also change in each sample (for a fixed choice of  $X$ ). This is because the functional form of  $H_X(\tau)$  makes use for example of the mean degree or other characteristics of the network. This means, that even if the distributions of  $\mathbb{H}^A$  and  $\mathbb{H}^B$  do not overlap, for a given  $X$  it is possible that the distributions of  $\tau_X^A$  and  $\tau_X^B$  do overlap. Fig. S19 gives a visual representation of this idea.

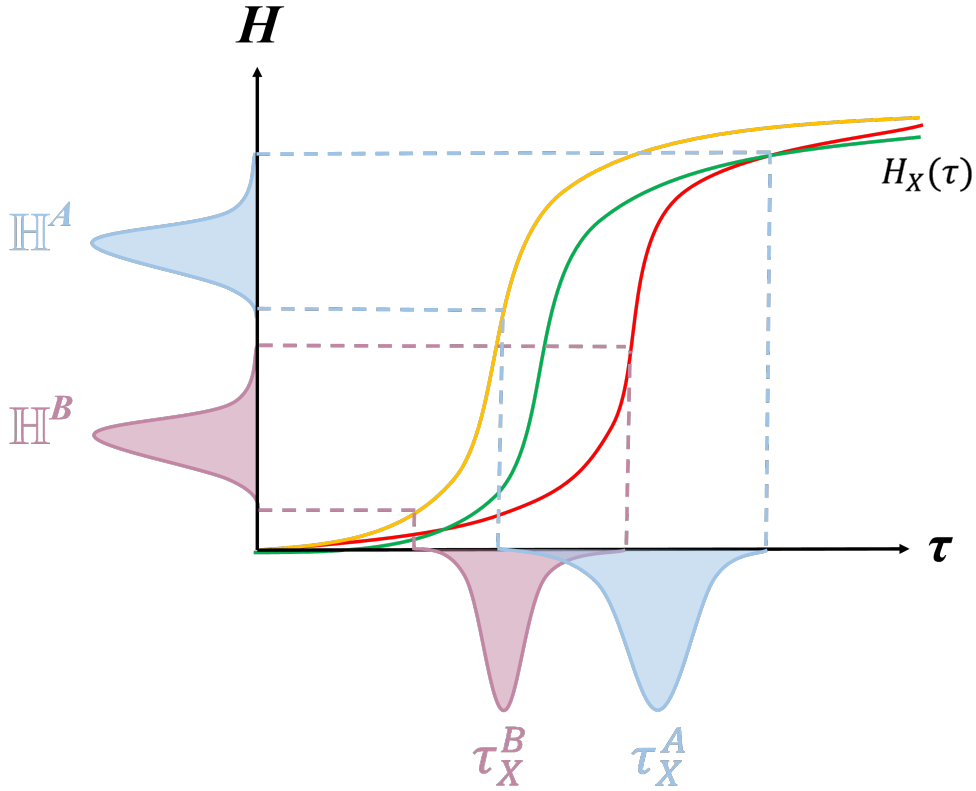


FIG. S19. Pictorial demonstration of how non-overlapping  $\mathbb{H}$  distributions from two traits  $A$  and  $B$  can lead to  $\tau_X$  distributions which do overlap through the bootstrapping process (see text). The yellow, green and red curves are examples of  $H_X(\tau)$  from different bootstrap samples.



### F. SUCRA scores and comparison with Dediu

To quantify the rankograms for each of the 35 linguistic traits shown in Fig. S18 further we use a statistic called SUCRA (surface under the cumulative ranking line) first introduced in [S34] to rank medical treatments. This is a transform of the expected rank,

$$\text{SUCRA}_a = \frac{N - \mathbb{E}(r)_a}{N - 1}, \quad (\text{S272})$$

where  $\mathbb{E}(r)_a$  is the expected rank of item  $a$  and where we are ranking  $N$  items total. SUCRA values go from 0 to 1 but are typically written as a percentage. If item  $a$  ranks first with probability 1 then it will have  $\text{SUCRA}_a = 1$  (or 100%). Similarly if item  $a$  ranks last with probability 1 then it will have  $\text{SUCRA}_a = 0$  (or 0%).

In Tab. S1 we list the SUCRA scores for all 35 traits for the different choices of the function  $H_X(\tau)$  as well as the SUCRA derived from a ranking based on  $\mathbb{H} = \sigma/[2\rho(1 - \rho)]$  itself. We also list Dediu's PC1 values where possible [S30]. We can calculate the Spearman rank correlation coefficient,  $r_S$ , between the traits sorted by SUCRA value and PC1. All give  $r_S > 0.8$  with  $p\text{-value} < 0.01$  after removing the outliers.

WALS ID	SUCRA values (%)					Dediu's PC1
	$\mathbb{H}$	$H_{AA}(\tau)$	$H_{SL}(\tau)$	$H_{HG}(\tau)$	$H_{SPA}(\tau)$	
130A	96.97	96.97	97.02	96.90	97.03	97.03
92A	94.52	94.36	94.33	94.60	94.44	87.10
11A	88.42	88.07	88.43	87.67	88.12	-101.68
143F	88.02	87.99	87.87	87.38	88.01	
143E	85.30	85.61	85.42	84.97	85.54	
33A	79.73	79.70	79.66	79.57	79.66	
38A	79.49	79.12	79.23	79.56	79.34	76.40
37A	78.77	78.77	78.64	78.77	78.77	87.38
8A	77.89	77.25	77.42	77.93	77.04	-57.03
77A	73.38	73.26	73.10	73.01	73.01	50.18
69A	72.26	72.28	72.26	72.26	72.26	
4A	66.43	66.44	66.52	66.43	66.40	-16.94
26A	61.30	61.49	61.35	62.21	61.60	
100A	58.05	57.77	57.63	57.21	57.47	
57A	55.69	55.90	55.76	55.51	55.83	40.60
6A	52.71	52.75	52.96	52.70	52.77	-68.03
7A	50.79	51.04	50.97	51.34	51.14	-42.87
129A	50.69	50.65	50.68	50.82	50.62	
120A	45.45	45.55	45.79	45.46	45.80	
48A	44.79	45.16	44.99	45.16	45.32	18.44
44A	44.47	44.48	44.68	44.08	44.27	-8.95
107A	41.41	41.64	41.23	41.44	41.50	79.84
70A	40.29	40.27	40.03	40.48	40.17	-23.10
87A	33.27	33.33	33.30	33.65	33.23	-11.23
73A	31.48	31.71	31.86	31.67	31.86	-64.56
119A	31.42	31.37	31.35	31.98	31.21	
82A	25.51	25.54	25.47	25.85	25.51	-69.15
13A	17.81	18.01	18.04	17.98	18.04	-35.89
9A	17.76	17.80	17.80	18.05	18.05	-20.70
91A	17.15	17.28	17.28	17.69	17.28	-23.96
27A	16.65	16.94	16.85	16.69	16.85	
83A	12.20	12.27	12.15	12.07	12.20	-28.65
89A	11.62	11.59	11.65	11.81	11.54	-30.62
86A	6.82	6.99	6.87	7.07	6.99	-60.98
53A	1.23	1.19	1.10	1.29	1.13	-43.21

TABLE S1. Table showing the SUCRA values for each of the 35 rankograms in Fig. S18 for all theories. Dediu's PC1 values are also shown where possible. The Spearman rank correlation coefficients,  $r_S$ , between each column and Dediu's PC1 is 0.824, except  $H_{SPA}(\tau)$  which is 0.832, all with  $p\text{-value} < 0.01$ .

To determine the outliers we recursively prune the sample. At each step we remove the trait that would give the largest increase in  $r_S$ . We show this in Fig. S20. We find that the four traits 11A, 8A, 6A and 7A give

significant improvement when removed. This is a different ordering than in Fig. S17, here it seems that 6A and 7A are outliers whereas 107A and 57A are not. This is likely simply due to using different pruning methods. It is encouraging however that 11A and 8A are still outliers which is the same as before. In [S1] there are strong linguistics arguments as to why Dediu incorrectly classifies 11A.

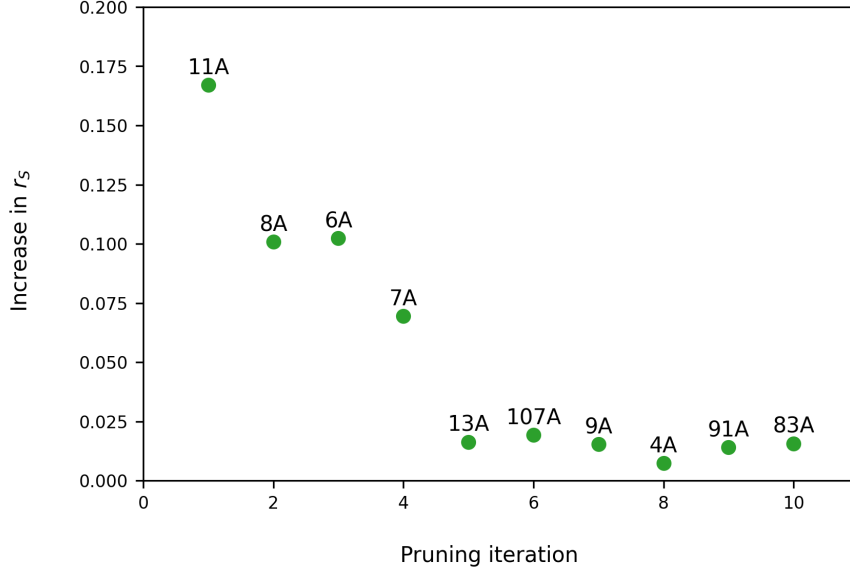


FIG. S20. Visual representation of outlier removal when determining the Spearman rank correlation coefficient,  $r_S$ , between the linguistic traits ranked by SUCRA value and Dediu's PC1 [see Tab. S1]. We recursively prune the data and remove the trait which would give the largest improvement in  $r_S$ .

## S10. Further mathematical background

### A. Exponential generating function for modified Bessel functions of the first kind

The exponential generating function for the modified Bessel function of the first kind is [S35, 49:6:1]

$$I_0(2x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(k!)^2} = \frac{x^0}{(0!)^2} + \frac{x^2}{(1!)^2} + \frac{x^4}{(2!)^2} + \dots \quad (\text{S273})$$

Squaring this equation provides us with the exponential generating function for  $I_0(2x)^2$ ,

$$\begin{aligned} I_0(2x)^2 &= \left( \frac{1}{(0!)^2} \frac{1}{(0!)^2} \right) x^0 + \left( 2 \frac{1}{(0!)^2} \frac{1}{(1!)^2} \right) x^2 \\ &\quad + \left( 2 \frac{1}{(0!)^2} \frac{1}{(2!)^2} + \frac{1}{(1!)^2} \frac{1}{(1!)^2} \right) x^4 + \dots \\ &= \sum_{k=0}^{\infty} \sum_{a_1+a_2=k} \frac{x^{2k}}{(a_1!)^2 (a_2!)^2}. \end{aligned} \quad (\text{S274})$$

To generalise, first define the combinatoric sum

$$\sum_{a_1+\dots+a_d=k} \equiv \sum_{A_d^k}. \quad (\text{S275})$$

In general then we can form the exponential generating function for  $I_0(2x)^d$ ,

$$\begin{aligned} I_0(2x)^d &= \sum_{k=0}^{\infty} \sum_{A_k^d} \frac{x^{2k}}{\prod_i^d (a_i!)^2} \\ &= \sum_{k=0}^{\infty} \sum_{A_k^d} \frac{(2k)!}{\prod_i^d (a_i!)^2} \frac{x^{2k}}{(2k)!}. \end{aligned} \quad (\text{S276})$$

Now recall the expression for  $W_d^{(2k)}$  from equation (S191),

$$W_d^{(2k)} = \sum_{A_k^d} \frac{(2k)!}{\prod_{i=1}^d (a_i!)^2}, \quad (\text{S277})$$

which gives us the final expression

$$I_0(2x)^d = \sum_{k=0}^{\infty} W_d^{(2k)} \frac{x^{2k}}{(2k)!}. \quad (\text{S278})$$

### B. Determining standard model parameters for a given $\tau$ and $\langle m \rangle_{\text{st}}$

In this section we describe the algorithm we used to generate model parameters in the simulation. The challenge is to generate sets of non-negative  $h_{\pm}, h'_{\pm}, v'_{\pm}$  leading to different stationary trait frequencies  $\langle \rho \rangle_{\text{st}}$  but at the same time keeping  $\tau = 2(h'_+ + h'_- + v'_+ + v'_-) / [(h_+ + h_-) - (h'_+ + h'_-)]$  fixed. Fixing  $\langle \rho \rangle_{\text{st}}$  is equivalent to fixing the stationary magnetisation  $\langle m \rangle_{\text{st}} = 2 \langle \rho \rangle_{\text{st}} - 1$ .

We apply the restriction in Eq. (S5), which we write in the form

$$h_+ = h'_+ - h'_- + h_-. \quad (\text{S279})$$

Equations (S24) and (S46) can be solved for  $v'_+$  and  $v'_-$ , assuming  $\tau \neq 0$

$$v'_+ = \frac{\tau}{2}(1 + \langle m \rangle_{\text{st}})(h_- - h'_-) - h'_+, \quad (\text{S280a})$$

$$v'_- = \frac{\tau}{2}(1 - \langle m \rangle_{\text{st}})(h_- - h'_-) - h'_-. \quad (\text{S280b})$$

#### 1. The case $\tau > 0$

For  $\tau > 0$  and  $-1 < \langle m \rangle_{\text{st}} < 1$ , we set  $h'_+$  and  $h'_-$  to small random values and then choose a value of  $h_-$  large enough so that  $v'_+, v'_-$  and  $h_+$  are non-negative. This amounts to choosing  $h_-$  such that

$$h_- \geq \max \left( 0, h'_- - h'_+, h'_- + \frac{2}{\tau(1 + \langle m \rangle_{\text{st}})} h'_+, \left[ 1 + \frac{2}{\tau(1 - \langle m \rangle_{\text{st}})} \right] h'_- \right). \quad (\text{S281})$$

For the case  $\tau > 0$  and  $\langle m \rangle_{\text{st}} = -1$ , Eq. (S280a) means that we need to set  $v'_+ = h'_+ = 0$ . Eq. (S280b) then gives  $v'_- = \tau h_- - (1 + \tau)h'_-$ . So to ensure  $v'_- \geq 0$  we draw random values of  $h_-$  and  $h'_-$  such that  $h_- \geq (1 + \frac{1}{\tau})h'_-$ . This allows us to calculate  $v'_-$ . Finally  $h_+$  is determined from Eq. (S279).

For the case  $\tau > 0$  and  $\langle m \rangle_{\text{st}} = +1$ , Eq. (S280b) means that we should set  $v'_- = h'_- = 0$ . Eq. (S280a) then gives  $v'_+ = \tau h_- - h'_+$ . So to ensure  $v'_+ \geq 0$  we draw random values of  $h_-$  and  $h'_+$  such that  $h_- \geq \tau h'_+$ . This allows us to calculate  $v'_+$ . The coefficient  $h_+$  is again determined from Eq. (S279).

2. The case  $\tau = 0$  (standard voter model)

For  $\tau = 0$  Eqs. (S280a) and (S280b) dictate that we need to set the rates of all unfaithful events to zero ( $h'_+ = h'_- = v'_+ = v'_- = 0$ ). We then have  $h_+ = h_-$  from Eq. (S279). The model therefore reduces to the standard voter model. The common value of the rates  $h_{\pm}$  can be drawn at random (the only significance of the value it to set a time scale). We note however that in this case the expression for  $\langle m \rangle_{\text{st}}$ , Eq. (S24), is not defined. In the voter model average magnetisation is conserved in time, and every initial magnetisation is therefore a steady-state value ( $\langle m \rangle_{\text{st}} = \langle m \rangle_0$ ).

3. Negative values of  $\tau$  ( $\tau \leq -2$ )

As discussed in Sec. S8 the quantity  $\tau$  cannot take values in the interval  $-2 < \tau < 0$ , that is, the only negative values of  $\tau$  permitted by the model setup are those in the range  $\tau \leq -2$ . We can confirm this through numerical search for model parameters. More precisely, we search for physical combinations of  $(v'_+, v'_-, h_+, h_-, h'_+, h'_-)$  which satisfy the constraint in Eq. (S5) for a fixed  $\langle m \rangle_{\text{st}}$  [Eq. (S5)] and  $\tau$  [Eq. (S46)]. For  $-2 < \tau < 0$  the search returns no solutions (see the Mathematica notebook at the following GitHub repository [S24] for a demonstration).

Focusing on  $\tau \leq -2$ , a simple method to generate model parameters is to set  $h'_+ = h'_-$  and then draw this number at random. Consequently, from Eq. (S279) we have  $h_+ = h_-$ . Then, to ensure  $v'_+ \geq 0$  and  $v_- \geq 0$ , Eqs. (S280a) and (S280b) give the following bounds on  $h_-$ :

$$0 \leq h_- \leq \min \left( 1 + \frac{2}{\tau(1 + \langle m \rangle_{\text{st}})}, 1 + \frac{2}{\tau(1 - \langle m \rangle_{\text{st}})} \right) h'_-. \quad (\text{S282})$$

To ensure the upper bound here is at least zero we require

$$\tau \leq \min \left( \frac{-2}{1 + \langle m \rangle_{\text{st}}}, \frac{-2}{1 - \langle m \rangle_{\text{st}}} \right). \quad (\text{S283})$$

This means that for a given  $\tau \leq -2$ ,  $\langle m \rangle_{\text{st}}$  must be within the following bounds.

$$-1 - \frac{2}{\tau} \leq \langle m \rangle_{\text{st}} \leq 1 + \frac{2}{\tau}. \quad (\text{S284})$$

A visual representation of these bounds is shown in Fig. S21.

The inequalities in Eq. (S284) bound the possible values of  $\langle m \rangle_{\text{st}}$  that can be reached for a given  $\tau$  by model parameters *generated using our algorithm*. In particular we used the restriction  $h_+ = h_-$  to derive these bounds. We can therefore not exclude that models with  $h_+ \neq h_-$  can cover a wider range of  $\langle m \rangle_{\text{st}}$  for a given  $\tau$ . However, numerical searches (allowing for  $h_+ \neq h_-$ ) find that the property from Eq. (S284) seems to continue to hold [S24].

We here provide further intuitive support for these bounds. From Sec. S8 we know that  $\tau \rightarrow -2^-$  corresponds to increasing amounts of anti-correlation between the spin states of nearest-neighbour nodes. This requires an increasingly equal proportion of up and down spin nodes [see Fig. S13(e)]. At  $\tau = -2$  this requires  $\langle m \rangle_{\text{st}} = 0$ , and as  $\tau$  becomes more negative from there the range of possible  $\langle m \rangle_{\text{st}}$  values increases.

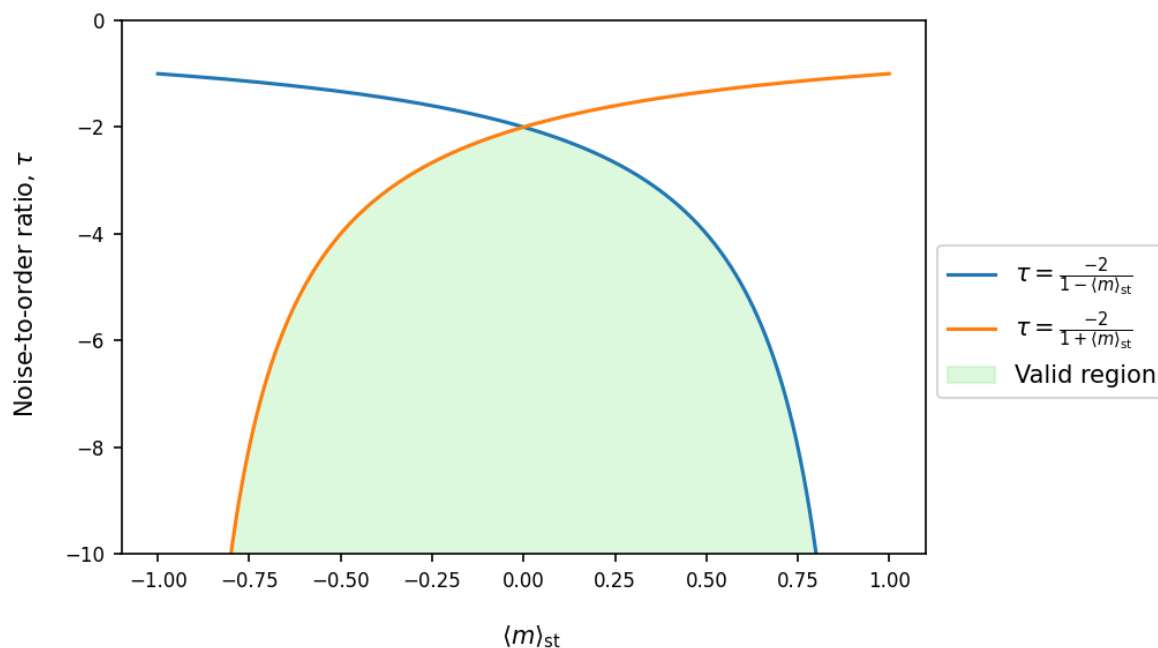


FIG. S21. Diagram showing the values of  $\tau \leq -2$  and  $\langle m \rangle_{st}$  for which it is possible to generate model parameters. The boundaries are given by Eq. (S283).

- 
- [S1] H. Kauhanen, D. Gopal, T. Galla, and R. Bermúdez-Otero, *Science Advances* **7**, eabe6540 (2021).
- [S2] F. Vazquez and V. M. Eguíluz, *New Journal of Physics* **10**, 063011 (2008).
- [S3] B. L. Granovsky and N. Madras, *Stochastic Processes and their Applications* **55**, 23 (1995).
- [S4] A. Carro, R. Toral, and M. S. Miguel, *Scientific Reports* **6**, 1 (2016).
- [S5] J. P. Gleeson, *Phys. Rev. X* **3**, 021004 (2013).
- [S6] S. Dorogovtsev, *Lectures on Complex Networks* (Oxford University Press, Oxford, UK, 2010).
- [S7] A. Knoblauch, *SIAM Journal on Applied Mathematics* **69**, 197 (2008).
- [S8] A.-L. Barabási, *Network science* (Cambridge University Press, Cambridge, UK, 2016) <https://networksciencebook.com/>.
- [S9] M. E. J. Newman, *Networks* (Oxford University Press, Oxford, UK, 2018).
- [S10] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*.
- [S11] R. L. Graham, D. E. Knuth, O. Patashnik, and S. Liu, *Computer in Physics* **3**, 106 (1989).
- [S12] P. Flajolet and R. Sedgewick, *Analytic combinatorics* (Cambridge University Press, 2009).
- [S13] M. D. Schmidt, *Axioms* **8**, 10.3390/axioms8020062 (2019).
- [S14] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and series volume 4: direct laplace transforms*, Vol. 4 (Gordon and Breach Science Publishers, Reading, UK, 1992).
- [S15] G. Lauricella, *Rendiconti del Circolo Matematico di Palermo* **7**, 111 (1893).
- [S16] K. Matsumoto, Appell and lauricella hypergeometric functions, in *Encyclopedia of Special Functions: The Askey-Bateman Project*, edited by T. H. Koornwinder and J. V. Stokman (Cambridge University Press, Cambridge, UK, 2020) p. 79–100.
- [S17] J. Abbott, *Laplace transforms of products of modified Bessel functions* (2022).
- [S18] D. E. Knuth, *The art of computer programming. Vol. 1: Fundamental algorithms* (Addison Wesley Longman, Boston, MA, 1997).
- [S19] J. M. Borwein and P. B. Borwein, *Pi and the AGM: a study in the analytic number theory and computational complexity* (John Wiley & Sons, Inc, Hoboken, NJ, USA, 1987).
- [S20] I. M. Wanless, *Combinatorics, Probability and Computing* **19**, 463 (2010).
- [S21] R. P. Stanley, *Catalan numbers* (Cambridge University Press, Cambridge, UK, 2015) pp. 1–14.
- [S22] K. Satō and Y. Iwasa, Pair approximations for lattice-based ecological models, in *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, Cambridge Studies in Adaptive Dynamics, edited by U. Dieckmann, R. Law, and J. A. J. Metz (Cambridge University Press, Cambridge, UK, 2000) pp. 341–358.
- [S23] A. F. Peralta, A. Carro, M. S. Miguel, and R. Toral, *New Journal of Physics* **20**, 103045 (2018).

- [S24] C. R. Kitching, [Estimating transmission noise on networks from stationary local order](#) (2024).
- [S25] C. W. Commander, Maximum cut problem, max-cut, in *Encyclopedia of Optimization*, edited by C. A. Floudas and P. M. Pardalos (Springer US, Boston, MA, 2009) pp. 1991–1999.
- [S26] R. Bermúdez-Otero, [Glossa: A Journal of General Linguistics](#) **5**, 1 (2020).
- [S27] R. Hickey, *The Handbook of Language Contact* (Blackwell Publishing Ltd, Hoboken, NJ, USA, 2010).
- [S28] M. S. Dryer and M. Haspelmath, eds., *WALS Online (v2020.4)* (Zenodo, 2013).
- [S29] B. Comrie, M. S. Dryer, D. Gil, and M. Haspelmath, in *The World Atlas of Language Structures Online* (Zenodo, 2013).
- [S30] D. Dediu, [Proceedings of the Royal Society B: Biological Sciences](#) **278**, 474 (2011).
- [S31] D. Dediu and M. Cysouw, [PloS ONE](#) **8**, e55009 (2013).
- [S32] S. Boslaugh, *Statistics in a nutshell: A desktop quick reference* (O’Reilly Media, Inc., Sebastopol, CA, USA, 2012).
- [S33] R. W. Johnson, [Teaching Statistics](#) **23**, 49 (2001).
- [S34] G. Salanti, A. E. Ades, and J. P. A. Ioannidis, [Journal of Clinical Epidemiology](#) **64**, 163 (2011).
- [S35] K. Oldham, J. Myland, and J. Spanier, *An atlas of functions* (Springer, New York, NY, 2008).