

PAPER • OPEN ACCESS

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

To cite this article: Xiang Li *et al* *J. Stat. Mech.* (2022) 013402

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

You may also like

- [Conditions for the classicality of the center of mass of many-particle quantum states](#)
Xavier Oriols and Albert Benseny
- [Plausible domain configurations and phase contents in two- and three-phase BaTiO₃-based lead-free ferroelectrics](#)
Vitaly Yu Topolov, Kumar Brajesh, Rajeev Ranjan et al.
- [STABLE CONIC-HELICAL ORBITS OF PLANETS AROUND BINARY STARS: ANALYTICAL RESULTS](#)
E. Oks

PAPER: Interdisciplinary statistical mechanics

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

Xiang Li¹, Mauro Mobilia¹, Alastair M Rucklidge¹ and R K P Zia^{2,3}

¹ Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom

² Center for Soft Matter and Biological Physics, Department of Physics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, United States of America

³ Physics Department, University of Houston, Houston, TX 77204, United States of America

E-mail: mmxl@leeds.ac.uk, m.mobilia@leeds.ac.uk, a.m.rucklidge@leeds.ac.uk and rkpzia@vt.edu

Received 30 July 2021

Accepted for publication 27 November 2021

Published 12 January 2022

Online at stacks.iop.org/JSTAT/2022/013402

<https://doi.org/10.1088/1742-5468/ac410f>




Abstract. We investigate the long-time properties of a dynamic, out-of-equilibrium network of individuals holding one of two opinions in a population consisting of two communities of different sizes. Here, while the agents' opinions are fixed, they have a preferred degree which leads them to endlessly create and delete links. Our evolving network is shaped by homophily/heterophily, a form of social interaction by which individuals tend to establish links with others having similar/dissimilar opinions. Using Monte Carlo simulations and a detailed mean-field analysis, we investigate how the sizes of the communities and the degree of homophily/heterophily affect the network structure. In particular, we show that when the network is subject to enough heterophily, an 'overwhelming transition'



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

occurs: individuals of the smaller community are overwhelmed by links from the larger group, and their mean degree greatly exceeds the preferred degree. This and related phenomena are characterized by the network's total and joint degree distributions, as well as the fraction of links across both communities and that of agents having fewer edges than the preferred degree. We use our mean-field theory to discuss the network's polarization when the group sizes and level of homophily vary.

Keywords: network dynamics, agent-based models, stochastic processes

 Supplementary material for this article is available [online](#)

Contents

1. Introduction	2
2. Model formulation and quantities of interest	4
2.1. Model update rules	4
2.2. Quantities of interest	6
3. Symmetric case, $m = 0$: summary of results	7
4. Asymmetric case, $m \neq 0$: simulation results	8
4.1. Fraction of adders and cross-community links	8
4.2. Mean degrees and associated variances	11
4.3. Overwhelming transition region characteristics	11
4.4. Degree distributions and joint degree distributions	12
4.5. Measures of polarization	16
5. Asymmetric systems: theoretical considerations	17
5.1. Framework for a general mean-field analysis	17
5.2. Systems with low asymmetry or $-J \ll 1$	18
5.2.1. Mean-field predictions of $\alpha_\sigma, \rho_\sigma, \mu_\sigma$ and p_σ	19
5.3. A refined mean-field theory for the degree distribution of the overwhelmed minority	20
5.4. Transition line	22
6. Conclusion and outlook	23
Acknowledgments	25
References	25

1. Introduction

The relevance of simple mathematical models in describing collective social phenomena has a long history [1–3], and the importance of relating ‘micro-level’ to ‘macro-level’ phenomena is well established [4–8]. Yet, it is only recently that connections between simple models of social dynamics and those traditionally used in statistical mechanics, such as Ising-like spin models, have been systematically exploited in ‘sociophysics’ or ‘opinion dynamics’ [9–14]. As a result, there is an intense field of research dedicated to the study of social networks by means of models and tools borrowed from statistical physics [15–18]. In particular, various dynamical processes have been studied on complex networks whose structure is random but static, see, e.g. [19–26], while in other models agents and links co-evolve [27–33].

Naturally, collective phenomena such as phase transitions and polarization that emerge from agent interactions have received great attention. An important example of social interaction is *homophily*, which is the tendency for nodes to create links with similar ones [7, 17, 34–37], while the tendency to establish ties with different others is referred to as *heterophily* [38–41]. Homophily is commonly seen in political parties [42–48], and the effects of this form of ‘assortative mixing’ on network dynamics have been investigated in sociological [8, 34, 35, 37, 49–52] and interdisciplinary physics studies [53–59]. In this literature, homophily is often modeled by assuming a biased probability of creating an edge or of rewiring an existing link (edge weighting), and notably features in growing [52, 55, 57, 60] and nodal attribute network models [53, 54, 61], and is often considered together with other ‘structural balance’ processes that mitigate tensions between connected agents [58, 61, 62].

While homophily appears to be ubiquitous in social networks with many examples of ‘birds of a feather flock together’ behaviors, heterophily appears to be more elusive, with some empirical evidence in team formation processes [63], and in professional cooperation networks [38, 40]. Quite interestingly however, in a two-community growing network according to preferential attachment, it has recently been found that heterophily is responsible for the increase of the average degree of the agents of the smaller group [55].

Here we consider an evolving network model in which links fluctuate continuously as the result of the homophilic/heterophilic interactions between individuals of two communities, holding one of two different opinions that remains fixed, who try to satisfy a prescribed *preferred degree* [64–66].

The objective of this work is to understand in some detail how homophily and heterophily affect the stationary state of the network and its emerging properties. While some of these aspects are considered in [67] for the special case of opinion groups of the same size, here we focus on the general case of communities of arbitrary sizes, showing that this leads to surprising results at a price of a considerably more challenging analysis. Our main contribution is the detailed characterization of the ‘overwhelming transition’ arising under enough heterophily in communities of different sizes. Remarkably, as observed in [67], the network then undergoes a transition separating a phase where it is homogeneous and an ‘overwhelming phase’ in which agents of the smaller community are overwhelmed by links from those of the larger group, with degrees greatly

exceeding the preferred degree. Here, we unveil the properties of the overwhelming phase (and ordinary regime) notably in terms of the total and joint degree distributions (JDDs) by devising suitable mean-field theories and Monte Carlo simulations.

The remainder of the paper is organized as follows: the formulation of the model and quantities of interest are introduced in the next section. In section 3, we summarize the main properties of the symmetric system reported in [67]. Our main findings regarding the ordinary and overwhelming phases are presented in sections 4 and 5 centred respectively on simulation results and mean-field analysis. Our conclusions are presented in section 6. Further details are provided in the supplementary material (<https://stacks.iop.org/JSTAT/2022/013402/mmedia>).

2. Model formulation and quantities of interest

Our model consists of N agents (nodes) with a varying number of connections (links) between them, forming a fluctuating network. Each agent $j = 1, \dots, N$ is endowed with one of two possible states (‘opinions’), $\sigma_j = \pm 1$, and a preferred number of links κ . A fraction n_{\pm} of the agents is in opinion state ± 1 , so that the network consists of a number $N_{\pm} = Nn_{\pm}$ of agents holding opinion ± 1 . Using the terminology of Ising-like models [68], (N_+, N_-) can be replaced by (N, m) , where $m \equiv (N_+ - N_-)/N = n_+ - n_-$ is an intensive quantity called ‘magnetization’.

The basis of our model is a preferred degree network (PDN) [64–66, 69] in which, at discrete time steps, different individuals (nodes) are chosen to add or cut links to others depending on whether its degree is less or greater than κ . The former are referred to as ‘adders’ and the latter as ‘cutters’. So as to avoid frozen networks, κ is chosen to be a half integer. Here, the PDN dynamics is supplemented by two ingredients: (i) the network consists of two communities of differing opinions and sizes; (ii) there are social interactions among agents embodied by the notion of *homophily*. Controlled by a parameter $J \in [-1, 1]$, homophily models the behavior of individuals who prefer to ‘make friends’ and establish links with those holding the same opinion (positive homophily or simply homophily, $J > 0$) or those with opposite opinions (negative homophily or simply heterophily, $J < 0$), see figure S1 and section S1 in the supplementary material.

By combining Monte Carlo and mean-field techniques, we here investigate how the network topology respond to parameters (κ, J, N, m) in its *out-of-equilibrium* stationary state (see section S2 of the supplementary material), and focus on the unexpected phenomena arising when $m \neq 0$, briefly reported in [67].

2.1. Model update rules

The dynamic rules of our model are conveniently stated by assuming discrete time steps, $t = 1, 2, \dots$. These rules are illustrated in figure 1. At each t , an agent i ($i = 1, \dots, N$) of degree $k_i \in [0, N - 1]$ is randomly chosen. For convenience, we will refer to nodes connected to i as its ‘neighbors’ and those not connected as ‘non-neighbors’. Then, after i is chosen,

- if $k_i > \kappa$, choose a neighbor j and delete the link

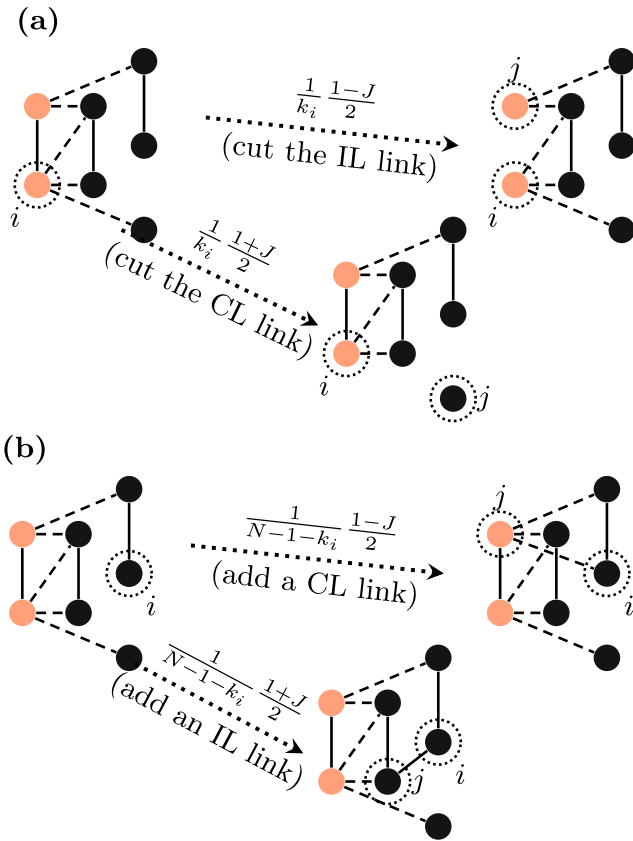


Figure 1. Illustration of the model for $\kappa = 2.5$. Dark nodes represent +1 voters (majority group) and light nodes are -1 voters (minority group). (a) Cutting process of an ij link when $k_i > \kappa$. (b) Adding of an ij link when $k_i < \kappa$. Dashed: cross links (CLs) between agents of different groups; solid: internal links (ILs) between voters of the same group. The probabilities of cutting an IL and adding a CL are $\tilde{J} = (1 - J)/2$ and $\hat{J} = (1 + J)/2$, respectively. Similarly, the respective probabilities of cutting a CL and adding an IL are \hat{J} and \tilde{J} , see text.

- * with probability \tilde{J} if $\sigma_i = \sigma_j$, or
- * with probability \hat{J} if $\sigma_i \neq \sigma_j$,
- if $k_i < \kappa$, choose one of the non-neighbors j and add a link
 - * with probability \hat{J} if $\sigma_i = \sigma_j$, or
 - * with probability \tilde{J} if $\sigma_i \neq \sigma_j$.

\hat{J} and \tilde{J} are defined by

$$\hat{J} \equiv (1 + J) / 2 \quad \text{and} \quad \tilde{J} \equiv (1 - J) / 2.$$

J thus plays a role similar to the nearest neighbor interaction in the ordinary Ising model (spin alignment). Note that for $J = 0$, the distinction between the communities is only nominal and as the opinions are irrelevant for the dynamics, and the system

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition becomes similar to the PDN models of [64, 69], see section S1 in the supplementary material.

2.2. Quantities of interest

To study the behavior of this evolving network in various regimes of parameter space, we focus on a few quantities of interest, summarized in the table of section S3 of the supplementary material. The most common of these is the degree distribution (DD) $p_\sigma(k)$ associated with agents in community σ , and from these p 's, we can extract the average degrees $\mu_\sigma \equiv \sum_k k p_\sigma(k)$ for each community $\sigma = \pm$, as well as the variances $V_\sigma \equiv \sum_k k^2 p_\sigma - \mu_\sigma^2$. We also distinguish the number of links an agent has to neighbors with opinion $\tau \in \{-, +\}$, and denote these by ℓ_τ (where $k = \ell_+ + \ell_-$). With ℓ_τ , we compile the JDDs [66] for the two communities: $P_\sigma(\ell_+, \ell_-)$, and from these can compute the averages:

$$(\bar{\ell}_\pm)_\sigma \equiv \sum_{\ell_+, \ell_-} \ell_\pm P_\sigma(\ell_+, \ell_-), \tag{1}$$

which can be regarded as the ‘centers of mass’ (CM) of the two JDDs, see figure 9 of [67]. The average degree in community σ is thus $\mu_\sigma = (\bar{\ell}_+)_\sigma + (\bar{\ell}_-)_\sigma$. Another convenient characteristic is the conditional distribution of cross-links:

$$q_\sigma(w | k) \equiv \frac{P_\sigma(\ell_+, \ell_-)}{p_\sigma(k)}; \quad w = \ell_{-\sigma}, \quad \ell_+ + \ell_- = k, \tag{2}$$

which gives the probability for a node in the group σ to have w cross-links (CLs), *provided* it has total degree k .

We also study the (fluctuating) total number of connections, which is a global quantity denoted by $L = L_\circ + L_\times$, where L_\circ is the overall number of internal links (ILs) and L_\times is the total number of CLs. Denoting by $L_{\sigma\tau}$ the links between agents opinions σ and τ , we have $L_\circ = L_{++} + L_{--}$ and $L_\times = L_{+-} = L_{-+}$, and hence

$$L = L_\circ + L_\times \equiv L_{++} + L_{--} + L_{+-}.$$

Since we focus on the steady state averages of these quantities, we simplify the notation by writing L in lieu of $\langle L \rangle$, etc. (The same below with α and ρ .) Clearly, these averages are related to $(\bar{\ell}_\tau)_\sigma$:

$$2L_{\sigma\sigma} = N_\sigma (\bar{\ell}_\sigma)_\sigma; \quad L_\times = N_+ (\bar{\ell}_-)_+ = N_- (\bar{\ell}_+)_-. \tag{3}$$

A natural way to describe polarization (extent of division between the communities) is to start with the ratio of CLs to the total number of links [32]

$$\rho \equiv L_\times / L, \tag{4}$$

and then a measure of polarization is given by

$$\Lambda \equiv 1 - 2\rho = (L_\circ - L_\times) / L, \tag{5}$$

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

so that $\Lambda(J = \pm 1) = \pm 1$. For asymmetric systems, however, the ratios for the two communities are distinct:

$$\rho_\sigma \equiv \frac{(\bar{\ell}_{-\sigma})_\sigma}{\mu_\sigma} = \frac{L_\times}{L_\times + 2L_{\sigma\sigma}}. \quad (6)$$

Further, as will be discussed in section 4.5 below, Λ suffers from some deficiencies. Instead, let us introduce an alternative measure of polarization, Π , which relies on the (normalized) *difference* between the two CMs,

$$\delta_x \equiv \frac{(\bar{\ell}_+)_{+} - (\bar{\ell}_+)_{-}}{(\bar{\ell}_+)_{+} + (\bar{\ell}_+)_{-}}, \quad \delta_y \equiv \frac{(\bar{\ell}_-)_{-} - (\bar{\ell}_-)_{+}}{(\bar{\ell}_-)_{-} + (\bar{\ell}_-)_{+}}.$$

Specifically, we define

$$\Pi \equiv \frac{\delta_x + \delta_y}{2} \in [-1, 1]. \quad (7)$$

In order to account for the (fluctuating) number of nodes to add/cut links, we denote by N^a and N^c the number of ‘adders’ and ‘cutters’, respectively. Further we denote by N_σ^β with $\sigma \in \{+, -\}$ and $\beta \in \{a, c\}$, the number of agents who are adders ($\beta = a$) or cutters ($\beta = c$) and hold opinion σ . Hence $N^a + N^c = N$ and $N_\sigma = N_\sigma^a + N_\sigma^c$, and the associated fractions are

$$n_\sigma^\beta \equiv N_\sigma^\beta / N; \quad n_\sigma = \sum_\beta n_\sigma^\beta; \quad n^\beta = \sum_\sigma n_\sigma^\beta. \quad (8)$$

Clearly, $\sum_{\sigma,\beta} n_\sigma^\beta = 1$. The fraction of adders, denoted by

$$\alpha \equiv n^a = N^a / N, \quad (9)$$

plays an important role. It is useful to define the fraction of adders in each group σ by

$$\alpha_\sigma \equiv N_\sigma^a / N_\sigma = n_\sigma^a / n_\sigma, \quad (10)$$

with $n_+ \alpha_+ + n_- \alpha_- = \alpha$ and $\alpha_\pm = \alpha$ when $m = 0$.

3. Symmetric case, $m = 0$: summary of results

For the sake of reference and completeness, we summarize the findings on the symmetric case $m = 0$ reported in [67]. We showed that by solving the balance equations for L_\times and L_\circ in the context of a mean-field approximation, the fractions of adders and CLs in the steady state of the symmetric case are

$$\alpha = \frac{1 - J^2}{2}, \quad \rho = \frac{1}{2} - \frac{J}{1 + J^2}, \quad (11)$$

which implies that the response to homophily in a clear unique mean-field expression of the polarization:

$$\Lambda = \Pi = 1 - 2\rho = \frac{2J}{1 + J^2}. \quad (12)$$

We also studied the stationary DDs $p(k)$ and $q(w|k)$. For the former, we found

$$p\left(\begin{matrix} k > \\ < \end{matrix} \kappa\right) = \left(\frac{1 \pm J^2}{3 \pm J^2}\right)^{\tilde{k}}; \quad \text{with } \tilde{k} = \begin{matrix} k - \lceil \kappa \rceil \\ \lfloor \kappa \rfloor - k \end{matrix}, \quad (13)$$

from which we obtain the degree average and its variance:

$$\mu = \mu_\sigma = \kappa + 3J^2/2, \quad V = V_\sigma = (7 + J^4)/4. \quad (14)$$

We also consider the JDD, $P_\sigma(\ell_+, \ell_-)$, with $P_+(x, y) = P_-(y, x)$ which, due to symmetry, gives the probability for either a + or - node to have x ILs and y CLs. The JDD is obtained from (2), where the conditional distribution of cross-links is approximated by a binomial distribution, i.e. $q \simeq q_{\text{bin}}(w|k) = \binom{k}{w} \rho^w (1 - \rho)^{k-w}$ [67]. Hence, with (13) and (11), the product

$$P_\sigma\left(\frac{k+u}{2}, \frac{k-u}{2}\right) \simeq p(k) q_{\text{bin}}((k - \sigma u)/2|k) \quad (15)$$

is a suitable approximation of the JDD for our purposes. As explained in section 4.4, for the generic $m \neq 0$ case with $|mJ| \ll 1$, the form (15) can be used to get a three-dimensional impression of the network's properties (see figure 7).

4. Asymmetric case, $m \neq 0$: simulation results

In our simulations, m varies from -0.04 to -0.6 , and we will see that $m = -0.04$ and $m = -0.6$ represent two very different regimes, i.e. the ordinary phase and the overwhelming phase, the system displays properties quite distinct from those summarized in section 3, see below.

We initiate the system with no initial links, run for typically over 10^2 Monte Carlo step (MCS), and verify that quantities like α_σ and ρ_σ have settled into steady values. One MCS is N updates. Thereafter, simulation runs are carried on for up to another 10^6 MCS, during which we measured various quantities every MCS. For simplicity, we show mainly the data associated with $N = 1000$ and $\kappa = 60.5$, so that $1 \ll \kappa \ll N_\pm$. The data collected for other values give some impression of finite-size effects, which can be serious when N and κ are lowered by an order of magnitude. In addition, we have carried out 10 runs with different random number generators in a handful of cases, to get a better idea of statistical errors. In all cases tested, the scatter for global quantities like α and ρ is much smaller than the size of the symbols (i.e. at most one part in a thousand).

4.1. Fraction of adders and cross-community links

As figure 2(a) illustrates, the asymmetric case is of great difference from the symmetric case even for the smallest m , except that asymmetry has no effect on the fraction of adders at $J = 0$. The effect of homophily appears to be *opposite* for the two α 's: when

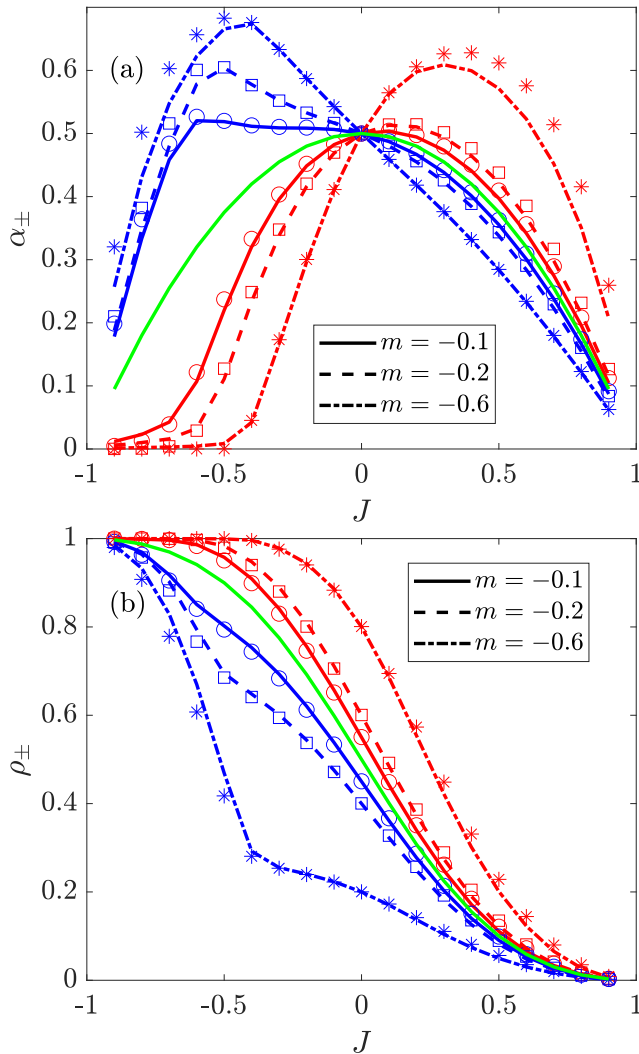


Figure 2. α_{\pm} and ρ_{\pm} vs J in the asymmetric case with different values of m with $N = 1000$ and $\kappa = 60.5$. Red lines and markers: agents of opinion +1 (minority); blue lines and markers: agents of opinion -1 (majority). Symbols \circ , \square and \star are associated with the simulation results when $m = -0.1, -0.2$ and -0.6 , respectively. Blue and red lines are mean-field predictions obtained by solving (16)–(18) and (26), see section 5. For comparison, green lines show the predictions (11) for $m = 0$.

$J > 0$ and $m < 0$, α_+ is greater than α in the symmetric case, while α_- is less than that α in the case $m = 0$. On the heterophily side ($J < 0$), we find richer phenomena, namely, the presence of a ‘kink’ in the $\alpha_-(J)$ curves at larger $-J$, accompanied by α_+ becoming vanishingly small. This phenomenon is most clearly seen in the $m = -0.6$ data, shown in figure 2(a), when $-J$ is larger than ~ 0.5 . The transition to this overwhelmed state is accompanied by rapid changes in the slopes of $\alpha_-(J)$. Such kinks can be seen even for low asymmetry systems (e.g. $m \simeq -0.1$), provided the strength of heterophily is sufficiently large, see section 4.3 As we probe deeper into this regime (i.e. larger $-J$), we find $\alpha_-(J)$ turning downwards, which is consistent with α vanishing in the limits $|J| = 1$.

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

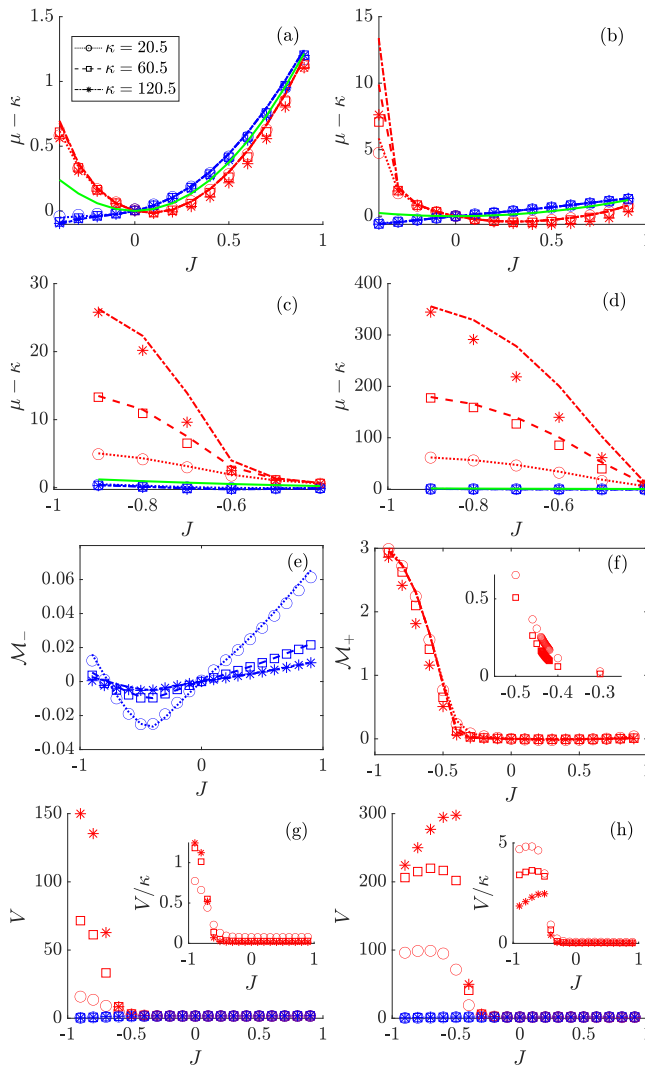


Figure 3. In all panels and insets, markers and lines denote $\kappa = 20.5$ (\circ , dotted), 60.5 (\square , dashed), and 120.5 ($*$, dash-dotted). Red and blue respectively refer to the minority (+) and majority (-) group. Symbols are from simulation data with $N = 1000$. In (a), (c) and (g), $m = -0.1$ while $m = -0.6$ in (b), (d)–(f) and (h). In (a)–(d), the green lines show the predictions (14) for $m = 0$. (a), (b) $\mu - \kappa$ vs J , under low (a) and high (b) asymmetry. Lines are solutions equations (16)–(18) and (26) of section 5. The extent of the J axis does not attain high heterophily. (c), (d) $\mu - \kappa$ vs J for $J \in [-1, -0.4]$ (significant heterophily). (e), (f) \mathcal{M}_{\pm} vs J for the majority (e) and minority (f). The inset of (f) shows a blow up of \mathcal{M}_+ vs J about transition region (more data points). (g), (h) DD variances V_- (blue) and V_+ (red) vs J under low (g) and high (h) asymmetry. Insets: V_{\pm}/κ vs J , see text.

For the fraction $\rho_\sigma(J)$ of CLs associated in the group holding opinion σ , we have $\rho_\pm(0) = n_\mp = (1 \mp m)/2$ and find a monotonic decrease with J in figure 2(b). Similar to the behavior of α_\pm , the curves of ρ_\pm vs J also deviate from the symmetric case in opposite directions. We find that the minority fraction, ρ_+ approaches 1, just as the majority ρ_- develops a kink, as J is decreased toward -1 . As may be expected, these features occur at about the same value of $J \approx J_c$ as those in $\alpha_\pm(J)$. Beyond the critical J , all ρ_\pm approach 1 rapidly when $J < J_c$, a property which can also be understood intuitively: heterophily is so strong in this regime that the entire population is locked into establishing CLs.

4.2. Mean degrees and associated variances

We now consider the mean degrees (μ_σ) and associated variances (V_σ) of each community $\sigma = \pm$. As in the case of α_\pm and ρ_\pm , the two μ 's deviate in opposite ways as we increase $|J|$, see figures 3(a) and (b). As the data for the larger κ 's show, the differences $\mu_\pm - \kappa$ in this regime converge on values which are $\mathcal{O}(1)$, which indicates that the communities are not interacting much. In stark contrast, interactions across the communities affect the network dramatically under larger heterophily, see figures 3(c) and (d): while μ_- remains relatively close to κ , $\mu_+ - \kappa$ in the minority group is strongly κ dependent, with a pronounced effect for large asymmetry, see figure 3(d) (note the scale of the $\mu - \kappa$ axis). In this regime, the minority starts being 'overwhelmed' once $-J$ rises beyond the aforementioned $-J_c$. Indeed, the average degree of minority agents can be *enhanced* by a factor $\mathcal{E} \equiv \mu_+/\kappa \simeq \check{J}N_-/N_+$ which can be much larger than unity [70]. Thus, instead of the difference $\mu_\sigma - \kappa$, we plot (for large asymmetry, $m = -0.6$)

$$\mathcal{M}_\sigma \equiv \frac{\mu_\sigma - \kappa}{\kappa}$$

vs J in figures 3(e) and (f). From simulation results for \mathcal{M}_- in figure 3(e), we conclude that $\mathcal{M}_- \rightarrow 0$ as $\kappa \rightarrow \infty$, consistently with $\mu_- - \kappa = \mathcal{O}(1)$. In figure 3(f), we plot $\mathcal{M}_+ = \mathcal{E} - 1$ vs J and, from the data collapse of the simulation results (red symbols), we conjecture that \mathcal{M}_+ converges to some definite, non-trivial thermodynamic limit. The behavior of \mathcal{M}_+ is hence reminiscent of the Ising magnetization, becoming non-zero below a critical temperature. In our simulation results, \mathcal{M}_+ appears to execute a smooth crossover through the transition region (inset of figure 3(f)). The DD variances paint a similar picture, with $V_\sigma = \mathcal{O}(1)$ in the ordinary regime and $V_+ = \mathcal{O}(\kappa)$ in the overwhelmed state, see figures 3(g) and (h). The overall behavior is qualitatively clear, but the dependence on the parameters is complex as illustrated in figure 3(h) (note the scale of the axis). In the inset of figure 3(h), V_+/κ scales, to some extent, with κ but without converging as $\kappa \rightarrow \infty$, which suggests the need for a further study of finite-size effects to draw conclusions about the thermodynamic limit.

4.3. Overwhelming transition region characteristics

Here, we highlight the transition region between the ordinary and overwhelming phases by providing a perspective of the simulation data based on the derivatives of α_- , ρ_- , μ_+ , V_+ in figure 4, see section S4 in the supplementary material. We report the

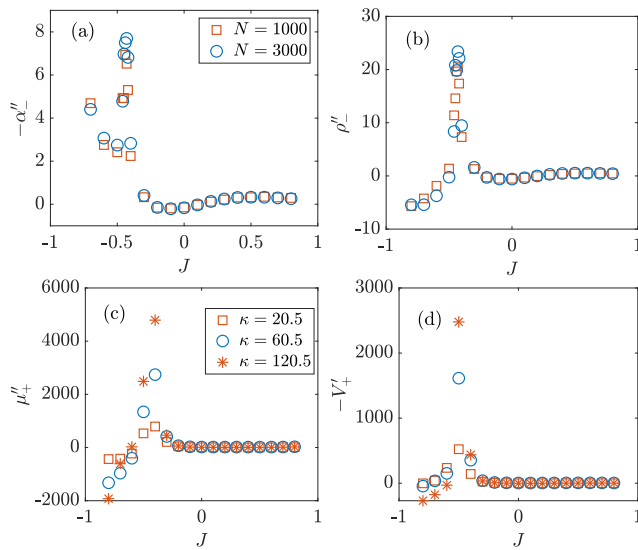


Figure 4. Highlights of the transition region for systems with $m = -0.6$ (where $' \equiv d/dJ$): (a) $-\alpha''$ vs J ; (b) ρ'' vs J ; (c) μ''_+ vs J ; (d) $-V'_+$ vs J . All results indicate a sharp peak around $J \approx -0.42$. Other parameters are: $N = 1000$ (\square), $N = 3000$ (\circ) and $\kappa = 60.5$ in (a) and (b); $N = 1000$, $\kappa = 20.5$ (\square), $\kappa = 60.5$ (\circ), and $\kappa = 120.5$ ($*$) in (c) and (d).

discrete derivatives of these quantities, and all results clearly indicate the existence of a sharp peak in the vicinity of $J \approx -0.42$. These peaks correspond to the ‘kinks’ shown in figure 2, at which the transition between ordinary and overwhelming phases occurs (‘overwhelming transition’). The critical point J_c at which this transition occurs depends on m , and $(m, J_c(m))$ is where $p(\lceil \kappa \rceil) \simeq p(\lceil \kappa \rceil + 1)$ at stationarity, see below and figure 6(d). Clearly, J_c is monotonically decreasing when m is negative, implying that the larger $|m|$, the less heterophily is needed to enter the overwhelming regime. When $m = 0$, the system falls into the ordinary regime for any value of J [67]. Remarkably, for any non-zero value of m (non-vanishing level of asymmetry), there is always a critical level of heterophily $|J_c|$, with $J_c < 0$, above which the system is in the overwhelming regime. The features of the transition line $(m, J_c(m))$ are well captured by the mean-field prediction given by equation (39), see section 5.4 and figure 5.

4.4. Degree distributions and joint degree distributions

We now study the total and JDDs. At low asymmetry and/or heterophily ($|m| \ll 1$ and/or $J > J_c$), the DDs $p_{\pm}(k)$ remain approximate exponentials (Laplacian distribution), dropping as k gets further from κ . As shown in figures 6(a)–(c), the Laplacian distributions are not symmetric, as the slopes on each side (log-linear plot) differ slightly, producing $\mu_{\sigma} \neq \kappa$, with the slopes of the DDs in the log-linear plots and widths being $\mathcal{O}(1)$. In our simulations, we found little dependence of the DDs on κ and N_{\pm} . The slopes of p_{\pm} differ in the opposite directions for the two communities, corresponding to α_{\pm} deviating from the $m = 0$ curve in opposite ways. In figure 6(c) we see that the

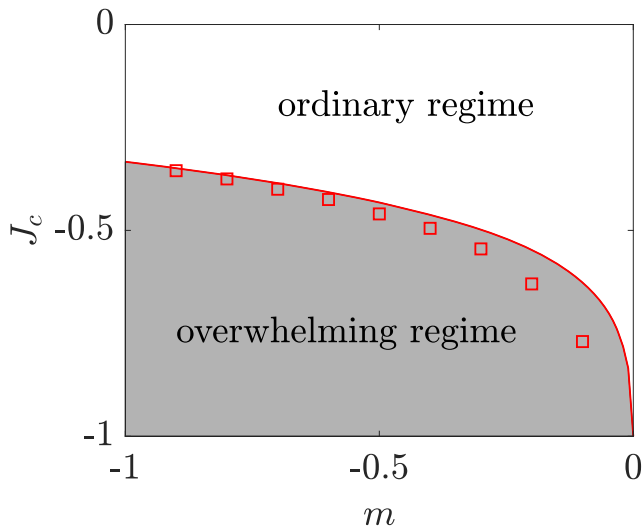


Figure 5. J_c vs m (with $N_+ < N_-$, i.e. $m < 0$) showing the transition line separating the ordinary (white) and overwhelming phases (shaded). Markers show (m, J_c) for which simulation data satisfying $p(\lceil \kappa \rceil) \simeq p(\lceil \kappa \rceil + 1)$. Here, $N = 1000$ and $\kappa = 60.5$; while the line is from equation (39). Note that a mirror diagram with $m \rightarrow -m$ is obtained in systems when $N_+ > N_-$ ($m > 0$).

right half of the DD of the minority is bent in the log-linear plot, and a Gaussian-like distribution for the cutters in the minority develops gradually as J is decreased further (increased heterophily), with J down to -0.6 in figures 6(d)–(f). These findings illustrate the process of the system transiting from the ordinary into the overwhelming regime. To locate the change of phase, the ‘overwhelming transition’, we assume that the system is in the transition regime when $p(\lceil \kappa \rceil) \simeq p(\lceil \kappa \rceil + 1)$, i.e. the slope of the DD at $\lceil \kappa \rceil$ of cutters in the minority equals 0, see figure 6(d). Within the overwhelming regime, the fraction of adders ($k < \kappa$) decreases substantially, and $p_+(k)$ from an exponential becomes a Gaussian-like distribution. In figures 6(e) and (f), we show the DDs deep in the overwhelming phase, where novel behavior emerges: while the majority keep their DD to be narrowly distributed around κ , the dramatic rise of the average degree of the minority agents is accompanied by significant changes to $p_+(k)$ and a substantial decrease of the adders in the minority. As illustrated in figure 6(f) for $m = J = -0.6$, there are no minority nodes with $k < \kappa = 60.5$ (i.e. no minority adders), while the distribution is essentially a Gaussian peaking at $k \simeq 145$, well over twice κ . Unlike the narrow Laplacian, the variance of this Gaussian is considerably higher, $V_+ \simeq 217$. The inset of figure 6(f) shows the striking difference between the DDs on linear scale.

We also consider the JDDs, $P_\sigma(\ell_+, \ell_-) = p_\sigma(k) q_\sigma(\ell_- | k)$ in the ordinary phase, see equation (2), where $k = \ell_+ + \ell_-$. In the regime where $|mJ|$ is small, we can approximate the conditional distribution of cross-links by a binomial distribution, i.e. $q_\sigma(\ell_- | k) \simeq \binom{k}{\ell_-} \rho_\sigma^{\ell_-} (1 - \rho_\sigma)^{k - \ell_-}$ [67] as in the symmetric case, yielding for the JDDs $P_\sigma(\ell_+, \ell_-) \simeq \binom{k}{\ell_-} p_\sigma(k) \rho_\sigma^{\ell_-} (1 - \rho_\sigma)^{k - \ell_-}$, that are of the same form as (15). Accordingly, the narrow Laplacian and broad Gaussian are embodied as different perspectives—in $p_\sigma(k)$ and

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

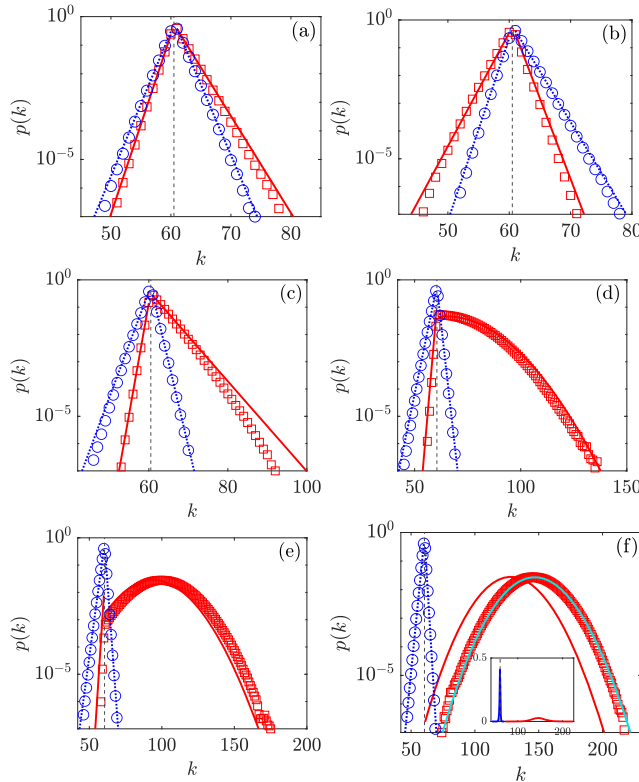


Figure 6. Total DDs $p_{\pm}(k)$ for various parameter sets in different regimes with $N = 1000$ and $\kappa = 60.5$. (a)–(c) p_{\pm} vs k in the ordinary regime; (a) $(m, J) = (-0.04, -0.5)$; (b) $(m, J) = (-0.6, 0.5)$; (c) $(m, J) = (-0.6, -0.3)$. (d) p_{\pm} vs k ($m, J) = (-0.6, -0.425)$, at onset of overwhelming transition where $p(\lceil \kappa \rceil) \simeq p(\lceil \kappa \rceil + 1)$, see text. (e), (f) p_{\pm} vs k in the overwhelming regime; (e) $(m, J) = (-0.6, -0.5)$; (f) $(m, J) = (-0.6, -0.6)$. Straight blue lines in (a)–(f) are from equation (23) given in section 5.2; the Gaussian-like curves in (d)–(g) are from equation (30), derived in section 5.4. Red curves in (d)–(f) are from equation (30) used together with equation (31), $\mu_-^{a,c} \approx \kappa$ and equation (34); the cyan curve in (f) is from equation (31) with α_- obtained from the simulation data (and $\mu_-^{a,c} \approx \kappa$).

$q_{\sigma}(\ell_{-\sigma}|k)$. A three-dimensional plot of $P_{\sigma}(\ell_+, \ell_-)$ is the best way to view the ‘knife-edge’ of the JDD caused by $p_{\sigma}(k)$ following a narrow asymmetric Laplacian distribution in the ordinary phase. As examples, in figure 7, we present the JDDs for the case of low asymmetry and intermediate heterophily, and the case of high asymmetry (and low heterophily).

In figures 7(a) and (b), the two perspectives for the *majority* JDD are shown, one along the knife-edge and the other, broadside. Note that $k = \ell_+ + \ell_-$, so that the perspective of the former is aligned with constant k . The latter clearly gives the impression of a Gaussian stemming from q_{σ} . This picture is qualitatively the same for the minority agents. In figures 7(c) and (d), we present the broadside perspective of the minority JDD in linear and semi-logarithmic scales. The latter reveals the initial stages of the minority being overwhelmed: in figure 7(d), the right side of the Gaussian is truncated near $\ell_+ = 0$, and squeezed toward $\ell_+ = 0$, implying that the probability of a minority

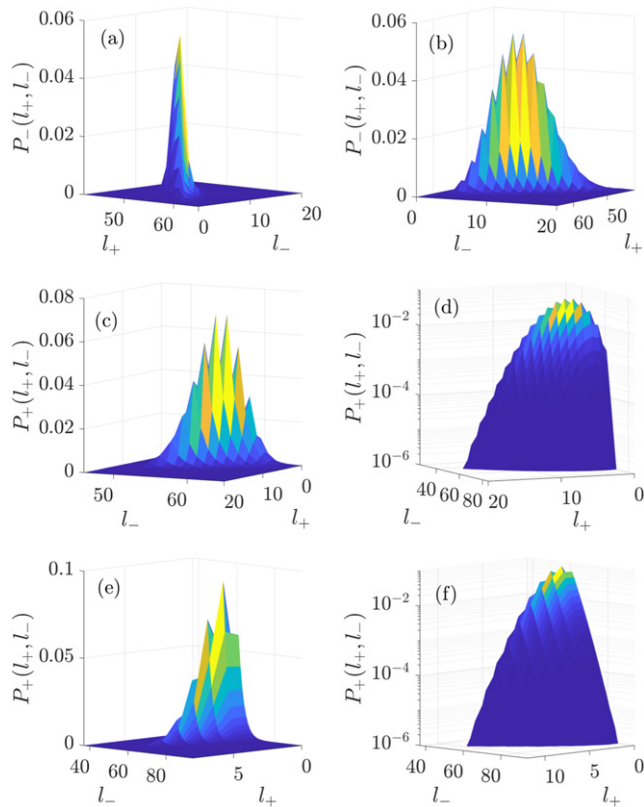


Figure 7. JDD $P_{\sigma}(\ell_{+}, \ell_{-})$ with $N = 1000$ and $\kappa = 60.5$ in the ordinary regime. Parameters are: $(N_{+}, N_{-}) = (480, 520)$, i.e. $m = -0.04$, and $J = -0.5$ in (a)–(d); $(N_{+}, N_{-}) = (200, 800)$, i.e. $m = -0.6$, and $J = -0.3$ in (e)–(f). (a), (b) Linear plots showing the ‘knife-edge’ (a) and broad Gaussian (b) perspectives of $P_{-}(\ell_{+}, \ell_{-})$. (c), (d) Linear (c) and semi log-scale (d) plots of $P_{+}(\ell_{+}, \ell_{-})$, where the minority JDD is pressed onto $\ell_{+} = 0$, see text. (e), (f) Linear (e) and semi-log (f) scale plots of $P_{+}(\ell_{+}, \ell_{-})$ under high asymmetry ($m = -0.6$) and low heterophily.

node to have a small number of ILs (small ℓ_{+}) is not vanishingly small. The graphs of figures 7(e) and (f), are similar, but for a low level of heterophily: the effect of heterophily is moderate, with almost half of the Gaussian being squeezed into the $\ell_{+} = 0$ plane, with the narrowing of the JDD along ℓ_{+} being accompanied by its broadening along ℓ_{-} . In this cross-over regime, the partition of $P_{+}(\ell_{+}, \ell_{-})$ into the product of a narrow Laplacian and a broad Gaussian is not valid (see figure 6(d)). Finally, deep in the overwhelming phase, the minority JDD collapses entirely onto the $\ell_{+} = 0$ plane, and the product expression is trivially valid: $P_{+}(\ell_{+}, \ell_{-}) = \delta_{\ell_{+}, 0} p_{+}(k = \ell_{-})$, where $p_{+}(k)$ is a broad Gaussian in the overwhelmed state. Meanwhile, the JDD of the majority agents, $P_{-}(\ell_{+}, \ell_{-})$, continues to display the same ‘knife-edge’ characteristics, as in the ordinary phase. In the transition region, the minority JDD cannot be simply approximated by the product of the DD and the conditional DD.

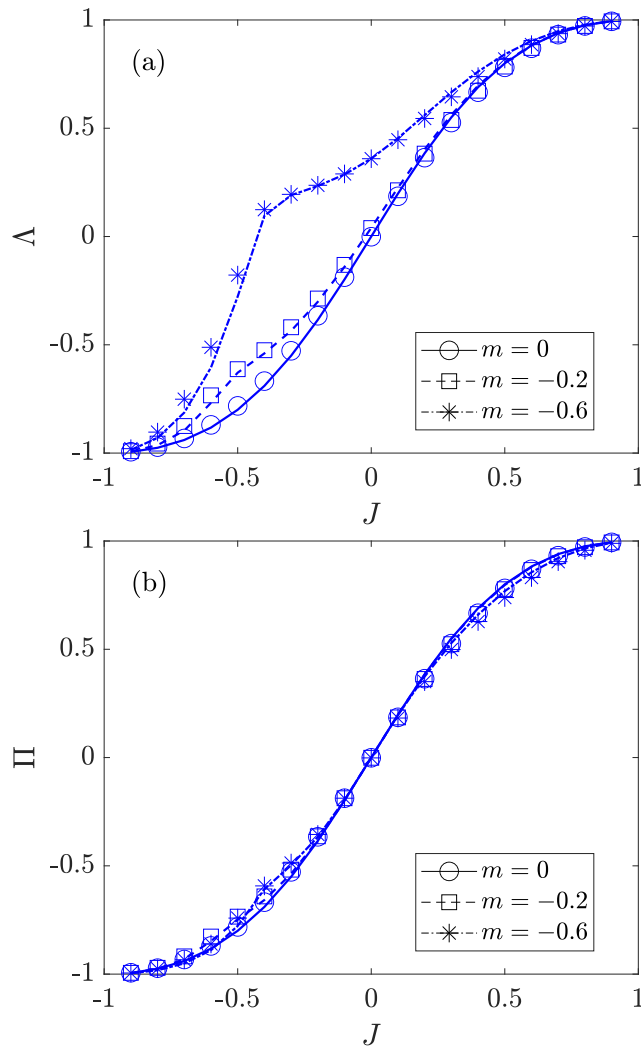


Figure 8. Polarization measures Λ and Π as functions of J for $m = 0$ (circles, solid), -0.2 (squares, dashed) and -0.6 (stars, dash-dotted); symbols are from simulation data and lines are mean-field predictions. Here, $N = 1000$ and $\kappa = 60.5$. (a) Λ vs J . (b) Π vs J . Lines in (a) are from equation (S6) and in (b) from equation (S8) of the supplementary material, with ρ_{\pm} obtained from the mean-field theory of section 5.2, see text.

4.5. Measures of polarization

When measuring the network polarization, the quantity Λ is commonly used and sometimes referred to as ‘average edge homogeneity’ [47, 71]. See section S6 in the supplementary material for the derivation of Λ in terms of ρ_{\pm} which are functions of J and m obtained from the mean-field theory of section 5.2. Λ provides a sensible measure of polarization in systems with low asymmetry ($m \approx 0$). However, this is generally not the case when $m \neq 0$, especially when $J \approx 0$ and $\Lambda \approx m^2$ fail to predict a vanishing polarization when $J \rightarrow 0$, see figure 8(a) and [67]. This led us to introduce Π , defined by

equation (7), which is constructed to avoid the deficiencies of Λ by providing a meaningful measure of polarization for any values of m . Although Π was motivated by the separation of the CMs of P_σ (the JDDs of the two groups), we have shown that in the context of the mean-field approximation, Π can be computed from ρ_\pm and m , see equation (S7) in the supplementary material. While Π contains the same information displayed in figure 2(b), its mean-field expression is an instructive combination of ρ_\pm offering a single meaningful quantity for polarization. We indeed find that $\Pi \propto 1 - \rho_+ - \rho_-$ which vanishes for $J = 0$ and any value of m (since $\rho_\pm = n_\mp$), while it reduces to Λ for $m = 0$. When $J = \pm 1$, $\Pi(\pm 1, m) = \Lambda(\pm 1, m) = \pm 1$. This implies that the sign of Π alone signifies if the system is homophilic or heterophilic, as seen in figure 8(b) where the mean-field predictions are in excellent agreement with simulation data (symbols) for all values of m and J . While Π does not appear to be independent of m , the dependence found in figure 8(b) turns out to be weak. Finally, we note that in figure 8, as expected, Λ and Π display a signal of the transition from the ordinary to the overwhelmed state about $J \approx J_c(m)$.

5. Asymmetric systems: theoretical considerations

From the phenomena presented in the last section, it is clear that there are two different regimes, the ordinary and overwhelming phases, with quite distinct properties. This section is devoted to their theoretical characterization in terms of mean-field theories. While we believe it is possible to formulate a single mean-field based theory which would completely describe both regimes, such a theory will be quite complex. Instead, here we adopt a simpler approach which has the advantage of being pragmatic and easily comprehensible, see section 5.2, at the price of being less effective in the overwhelming regime, an issue that we circumvent in section 5.3 with a refined (complementary) mean-field theory.

5.1. Framework for a general mean-field analysis

Our starting point to set a general framework for a mean-field analysis that applies for any choice of N_\pm , for arbitrary m , is to generalize the mean-field theory devised in [67] for the symmetric case $m = 0$, see also section 3.

For arbitrary m , we now have four unknowns: α_\pm (or n_\pm^a) and ρ_\pm [72]. As in the symmetric case we have two equations and still need two other equations to obtain a self-contained theory. One such equation is for each $L_{\sigma\sigma}$, while the other is a strict constraint involving ρ_\pm in the identity $L_{+-} = L_{-+}$.

Focusing first on global quantities like L , we must consider the changes for $L_{\sigma\sigma}$, with $\sigma = \pm$. Proceeding as in [67], the gain and loss rates are $n_\sigma^a n_\sigma \hat{J}$ and $n_\sigma^c (1 - \rho_\sigma) \hat{J}$, respectively. For the former, n_σ means that a node can choose to add a link to *unequal* fraction of partners [73], rather than 1/2 when $m = 0$. Now, there are four ways $L_{\sigma\sigma}$ can arrive at a stationary state. Two correspond to extreme values of J , when one of the rates vanishes and $L_{\sigma\sigma}$ reaches its own bound. Another is when both n_σ^a and $1 - \rho_\sigma$ vanish (or are vanishingly small), an interesting possibility we will return to in subsection 5.3. Here,

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition we are most interested in the last scenario, when both rates reach stationary, generic values. Balancing these, we find for the steady state: $\hat{J}n_\sigma^a n_\sigma = \check{J}n_\sigma^c (1 - \rho_\sigma)$, which is equivalent to

$$\alpha_\sigma n_\sigma B = (1 - \alpha_\sigma) (1 - \rho_\sigma), \tag{16}$$

where

$$B \equiv \hat{J}/\check{J} = \frac{1 + J}{1 - J}$$

is a convenient way to express the bias due to homophily: $B > 1$ for $J > 0$. For CLs, the generalization of the gain/loss rates is slightly more complicated, yielding

$$(n_+^a n_- + n_-^a n_+) \check{J} = (n_+^c \rho_+ + n_-^c \rho_-) \hat{J}, \tag{17}$$

where we can read the contributions from both groups.

Another equation comes from the constraint $L_{+-} = L_{-+}$. In terms of the variables here, this equality (trivially satisfied in the symmetric case) reads

$$N_+ \mu_+ \rho_+ = N_- \mu_- \rho_-. \tag{18}$$

In order to apply equation (18), we must first generalize the technique of [67] for the DDs $p_\pm(k)$ when $m \neq 0$, from which to extract μ_\pm in terms of α_\pm and ρ_\pm . Thus, we devote the next subsections to studies of the DDs.

5.2. Systems with low asymmetry or $-J \ll 1$

As seen in the simulation data, the DDs for asymmetric system with small m and $-J \ll 1$ are qualitatively the same as in the case $m = 0$. Proceeding as in the symmetric case [67], we start from the balance equation for the addition/deletion of a link at a single node:

$$R_\sigma^a(k) p_\sigma(k) = R_\sigma^c(k+1) p_\sigma(k+1), \tag{19}$$

for which we need expressions for the four $R_\sigma^{a,c}$ which are the probabilities in a time unit at which a connection is added (R_σ^a) and cut (R_σ^c) in the community σ . Each R has contributions from both communities, and the probabilities for adding and cutting links are here associated with the symbols η and χ , respectively. We will further distinguish contributions due to the actions (η, χ) of the chosen node, or from the rest of the population $(\tilde{\eta}, \tilde{\chi})$ [67]. In the context of our mean-field theory, the former pair simply reads:

$$\eta_\sigma = n_\sigma \hat{J} + n_{-\sigma} \check{J}, \quad \chi_\sigma = (1 - \rho_\sigma) \check{J} + \rho_\sigma \hat{J} \tag{20}$$

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition

as the prefactors of the J 's just refer to the chances our node adds/cuts a link to an agent in its community or otherwise. The latter is similar, except for the extra factors accounting for the fraction of adders/cutters in the σ group, yielding

$$\begin{aligned}\tilde{\eta}_\sigma &= \alpha_\sigma n_\sigma \hat{J} + \alpha_{-\sigma} n_{-\sigma} \check{J} \\ \tilde{\chi}_\sigma &= (1 - \alpha_\sigma) (1 - \rho_\sigma) \check{J} + (1 - \alpha_{-\sigma}) \rho_{-\sigma} (n_{-\sigma}/n_\sigma) \hat{J}.\end{aligned}\tag{21}$$

These enter into the R 's as in the case $m = 0$ [67]:

$$R_\sigma^a = \frac{[H(\kappa - k)\eta_\sigma + \tilde{\eta}_\sigma]}{N}; \quad R_\sigma^c = \frac{[H(k - \kappa)\chi_\sigma + \tilde{\chi}_\sigma]}{N}.\tag{22}$$

Proceeding as in the symmetric case, solving the recursion relation (19) [67], we obtain again asymmetric Laplacian distributions (see section S5 in the supplementary material and (13)):

$$\begin{aligned}p_\sigma(k < \kappa) &= p_\sigma(\lfloor \kappa \rfloor) [\gamma_{\sigma <}]^{\lfloor \kappa \rfloor - k} \\ p_\sigma(k > \kappa) &= p_\sigma(\lceil \kappa \rceil) [\gamma_{\sigma >}]^{k - \lceil \kappa \rceil},\end{aligned}\tag{23}$$

where, with (20) and (21),

$$\gamma_{\sigma <} = \frac{\tilde{\chi}_\sigma}{\eta_\sigma + \tilde{\eta}_\sigma}; \quad \gamma_{\sigma >} = \frac{\tilde{\eta}_\sigma}{\chi_\sigma + \tilde{\chi}_\sigma}\tag{24}$$

are the factors controlling the exponentials. Finally, by imposing the normalization conditions

$$\frac{p_\sigma(\lfloor \kappa \rfloor)}{1 - \gamma_{\sigma <}} = \alpha_\pm, \quad \frac{p_\sigma(\lceil \kappa \rceil)}{1 - \gamma_{\sigma >}} = 1 - \alpha_\pm,\tag{25}$$

both p 's are uniquely determined in terms of the unknowns α_\pm and ρ_\pm .

Our goal is to find the averages μ_σ of these distributions, which allow us to apply equation (18) and in turn complete our mean-field theory. From equations (23) and (25), we have

$$\mu_\sigma = \lfloor \kappa \rfloor - \frac{\gamma_{\sigma <} \alpha_\sigma}{1 - \gamma_{\sigma <}} + \frac{1 - \alpha_\sigma}{1 - \gamma_{\sigma >}},\tag{26}$$

where we have neglected terms of order $\mathcal{O}(\gamma_{\sigma >}^{\lfloor \kappa \rfloor}, \gamma_{\sigma <}^{\lceil \kappa \rceil})$ since $\kappa \gg 1$. Equation (26) gives us the expression of μ_\pm in terms of α_\pm and ρ_\pm , which can be determined by solving equations (16)–(18), and in turn allow us to obtain the predictions of this theory. Yet, since equations (16) and (18) are non-linear, we do not have explicit solutions. Instead we have solved equations (16)–(18) and (26) numerically by standard methods.

5.2.1. Mean-field predictions of $\alpha_\sigma, \rho_\sigma, \mu_\sigma$ and p_σ . The theoretical results based on equations (23), (16), (17), (18) and (26) have been directly used to obtain the mean-field predictions of α_\pm and ρ_\pm , Λ and Π (see section S6 in the supplementary material), μ_\pm and $\mathcal{M}_\pm = (\mu_\pm/\kappa) - 1$, and also the degree distribution $p_\sigma(k)$.

The blue and red lines of figure 2 show the predictions of α_σ and ρ_σ which are found in good accord with simulation data for all values of J , across both the ordinary and

overwhelming phases. The agreement is excellent in the ordinary regime, and we note that our mean-field theory remarkably captures the ‘kinks’ of α_- and ρ_- at the onset of the overwhelming transition. In the overwhelming phase, $|m| = \mathcal{O}(1)$ and $J < J_c < 0$ (see section 4.3), the degrees of the minority agents are no longer small compared to N , which violates a key assumption of our mean-field theory. This leads to systematic but rather modest deviations between the mean-field predictions and simulation data.

The comparison of simulation and theoretical results for Λ and Π in figure 8 shows a remarkable agreement for different values of m over the entire range of J , i.e. across both ordinary and overwhelming phases. In particular, the mean-field theory correctly captures the weak m -dependence of Π , and signals the transition from the ordinary to overwhelming regimes for both Λ and Π .

For μ_{\pm} and \mathcal{M}_{\pm} , as shown in figure 3, mean-field predictions generally agree well with simulation results, with an agreement that improves as $\kappa \ll N$, with $\kappa \gg 1$. Remarkably, the mean-field results give sensible results for μ_{\pm} , and \mathcal{M}_{\pm} also in the overwhelming regime, see section 5.3 and section S8 in the supplementary material.

Theoretical predictions of $p_{\sigma}(k)$ are used to obtain the red and blue lines in figure 6, which are generally in good agreement with simulation data, especially in the ordinary regime, i.e. $|mJ| \ll 1$ or $J > 0$, see figures 6(a) and (b). As shown in figure 6(c), for $|m| = \mathcal{O}(1)$ and above a certain level of heterophily ($J < 0$ with $|J| = \mathcal{O}(1)$), a Gaussian-like distribution for the cutters in the minority begins to develop, which is not captured by the above mean-field theory. In fact, deep in the overwhelming phase the DD of the minority no longer falls off exponentially as predicted by (23), but is a broad Gaussian, see figure 6(f), characterized in section 5.3. We note that the DD of the majority always falls off exponentially and, in both ordinary and overwhelming phases, is well described by (23), see blue lines in figures 6(a)–(f).

5.3. A refined mean-field theory for the degree distribution of the overwhelmed minority

Clearly, the overwhelmed states lie beyond the domain of validity of the above mean-field theory. This chiefly results from the DDs of the minority agents having morphed, see figures 6(e) and (f). To get a reasonable characterization of the quantities in this region, we have to study the DD of the minority agents more carefully.

For this, we revise the above theory following the approach of [66], and write the balance equation obeyed by the minority DD with a full non-trivial k -dependence of adding/cutting probabilities $R^{a,c}(k)$:

$$R_+^a(k-1)p_+(k-1) = p_+(k)R_+^c(k). \quad (27)$$

Furthermore, guided by simulation data, we assume that deep in the overwhelming phase we have $\alpha_+ = 0$ (see also at the end of this section). To determine $R^a(k)$, we recognize that there are just $(N_- - k)$ majority nodes which can add a link to our agent of degree k , each of which can be chosen with probability $1/N$, and only a fraction α_- of them would add. There is also the bias \check{J} for adding CLs. Finally, the adder will choose our agent with probability $1/(N - \check{k} - 1)$, where \check{k} is the number of links it has, which is a fluctuating quantity. In the spirit of a mean-field approximation, \check{k} is replaced by μ_-^a

Effects of homophily and heterophily on preferred-degree networks: mean-field analysis and overwhelming transition (average degree of majority adders). Putting everything together, we have

$$R_+^a(k-1) \simeq \frac{N_- - k + 1}{N} \frac{\alpha_-}{N - 1 - \mu_-^a} \tilde{J}. \quad (28)$$

Similar arguments give the expression of $R_+^c(k)$:

$$R_+^c(k) \simeq \frac{1}{N} \left(1 + k \frac{1 - \alpha_-}{\mu_-^c} \right) \hat{J}. \quad (29)$$

With (28) and (29), the solution of the recursion (27) is a ‘shifted binomial’ (see section S7 in the supplementary material):

$$p_+(k) \propto \frac{(Q^c/Q^a)^k (N_-)!}{\Gamma(Q^c + k + 1) (N_- - k)!}, \quad (30)$$

where

$$Q^a = (N - \mu_-^a - 1) B / \alpha_-; \quad Q^c = \mu_-^c / (1 - \alpha_-). \quad (31)$$

Since $Q^c > 0$, this expression is well-defined as far as $k = 0$. Since $\alpha_+ \approx 0$ is our assumption in the overwhelming state we expect this theory to be fairly good deep in the overwhelming regime, but to fail near the transition line.

For the characterization of the DD by (30), we expect $p_+(k)$ to approach a Gaussian in the limit of large N, κ with generic values of m, J . Thus, we use the mode \hat{k} for μ_+ and the curvature of $\ln p_+$ for V_+ , see also section S7 in the supplementary material

$$\mu_+ \simeq \hat{k} \simeq \frac{N_- + 1 - Q^a}{1 + Q^a/Q^c}. \quad (32)$$

$$V_+ = \frac{N_- Q^c Q^a}{(Q^a + Q^c)^2} + \frac{(Q^a)^2 + 2Q^a Q^c - (Q^c)^2 + 2Q^a (Q^c)^2}{2(Q^a + Q^c)^2}. \quad (33)$$

To determine Q^a and Q^c , guided by simulation data showing that the DDs for the majority agents still follow the asymmetric Laplacian distribution (23), we assume that $\mu_-^a \approx \mu_-^c \approx \kappa$. Also, consistently with our previous assumption, in the overwhelming state, we set $\alpha_+ = 0$ in (16) and (17), and obtain

$$\alpha_- = \frac{1 - J^2}{(1 - mJ)(1 - m)}. \quad (34)$$

These together give us the approximation of Q^a and Q^c in the overwhelming state, and thus the DD $p_+(k)$ of the minority agents (almost all ‘cutters’) shown as red curves in figures 6(d)–(f). Since (30) with (34) and $\mu_-^{a,c} \approx \kappa$ rely on the observed absence of adders and ILs in the minority community, it is a ‘semi-phenomenological’ refined mean-field theory. A fully self-consistent refined mean-field theory is outlined in section S9 of the supplementary material. Comparison with simulation results of figure 6 shows that this semi-phenomenological theory gives a good description of the minority DD not too deep in the overwhelming phase, see figures 6(d) and (e). Yet, deep in the overwhelming

regime of very heterophilic systems ($-J$ close to 1), there are quantitative deviations between the theoretical predictions and simulation data, see the red curve in figure 6(f). These are traced back to the use of (34), and a significantly better agreement is found when (30) is used with α_- directly obtained from simulations (and $\mu_-^{a,c} \approx \kappa$), leading to the cyan curve in figure 6(f). The accuracy of the predictions of (30) with (34) improves as the system size is increased, i.e. the red and cyan curves of figure 6(f) will get closer for larger N and κ .

As a simple assessment of our refined mean-field theory, we compare its predictions with the simulation data in the case of $N = 1000$ and $\kappa = 60.5$, with $m = J = -0.6$. As shown in figure 6(f), $p_+(k)$ in this case study is clearly Gaussian-like, with measured $\mu_+ \simeq 146$ and $V_+ \simeq 217$. These values are compared with the predictions of our theory based on equations (31), (30) and (34), yielding $(\mu_+, V_+) \simeq (128, 203)$ from (32) and (33). These results are in reasonable but not perfect agreement with those of simulation. The data can also be compared with (32) and (33) when these are used with α_- directly measured from simulations, yielding $(\mu_+, V_+) \simeq (146, 216)$, which are the approximation of the mean and variance of the cyan curve and compare remarkably well with those obtained from simulations. This agreement gives us confidence that we have devised a suitable mean-field description of the DD of the minority agents.

We conclude that our results, illustrated by figures 2–8, show that the ordinary MF approximation gives a sound qualitative and quantitative characterization of all quantities in the ordinary phase, as well as of the global quantities in the overwhelming phase and DDs of the majority phase. Yet, the refined MF is necessary to describe the DD of the minority in the overwhelming phase, see also section S8 and figure S2 in the supplementary material.

5.4. Transition line

In this section, we use the theoretical results of sections 5.2 and 5.3 to derive the mean-field prediction of the transition line $(m, J_c(m))$ separating the ordinary and overwhelming phases (respectively at $J > J_c$ and $J < J_c$, with m fixed), see figure 5. For the sake of concreteness, and without loss of generality, here we consider $m < 0$.

To find the point where the transition occurs, we start from the balance equation for a minority node of degree $k = \lceil \kappa \rceil$ which, from (19), reads

$$R_+^a(\lceil k \rceil)p_+(\lceil k \rceil) = p_+(\lceil k \rceil + 1)R_+^c(\lceil k \rceil + 1), \quad (35)$$

with

$$R_+^a(\lceil k \rceil) = \tilde{\eta}_+/N, \quad R_+^c(\lceil k \rceil) = (\chi_+ + \tilde{\chi}_+)/N, \quad (36)$$

where we have used (22) with $H(\kappa - \lceil \kappa \rceil) = 0$ and $H(\lceil \kappa \rceil - \kappa) = 1$. As discussed in section 4.3, we consider that the transition between the ordinary and overwhelming phases occurs when $p_+(\lceil k \rceil) = p_+(\lceil k \rceil + 1)$, see figure 6(d). With (35) and (36), this readily gives

$$\tilde{\eta}_+ = \chi_+ + \tilde{\chi}_+. \quad (37)$$

Further, we assume that at the onset of the transition, the features of both phases hold: $\alpha_+ = 0$, $\rho_+ = 1$ (see figure 2), and $\mu_+ \approx \mu_-$. With these assumptions and equations (3), (6), (20), (21) and (37), we have

$$\alpha_- n_- \check{J} = \hat{J} + (1 - \alpha_-) \hat{J}, \quad (38)$$

where $n_- = (1 - m)/2$ and $\alpha_-(J, m)$ can be approximated by equation (34). The unique physical root of (38) thus gives us the mean-field expression of $J_c(m)$, which explicitly reads (for $m < 0$):

$$J_c = \frac{\left(3m - 1 - 2m^2 + 2\sqrt{m(m^3 - 3m^2 + 4m - 1)}\right)}{1 + m}. \quad (39)$$

This expression is plotted in figure 5. The predictions of (39) are found to generally agree well with simulation data, with an excellent agreement for $m \lesssim -0.4$ and some noticeable deviations close to the symmetric case ($|m| \ll 1$). These are due to the deterioration of the approximation of $(\alpha_+, \rho_+) \approx (0, 1)$ that we attribute chiefly to finite size effects, expected to be important close to the ordinary phase consisting of a finite fraction of adders and CLs (given by (11) in both communities). Naturally, this and the limited validity of the crude assumption $\mu_+ \approx \mu_-$ affect the applicability of (39) [74].

6. Conclusion and outlook

We have investigated a dynamic, out-of-equilibrium network of individuals that may hold one of two different ‘opinions’ in a two-party society. In this work, the opinions of agents are held fixed while inter-party and cross-party links are endlessly created and deleted in order to satisfy a preferred degree. The evolving network has therefore a fluctuating number of links and is shaped by homophily and heterophily which model forms of social interactions by which agents tend to establish links with others having similar or dissimilar opinion, respectively. In our model, homophily/heterophily is modeled by an evolutionary process leading to the continuous ‘birth’ and ‘death’ of links within and between the communities. While the features of the system where the two opinion groups are of the same size (symmetric case) have been studied elsewhere [67], here we have focused on the generic case of communities of different sizes. We have thus investigated how the joint effect of community size asymmetry and homophily/heterophily influences the network structure in its steady state and leads to new phenomena.

The most striking feature of our model is the transition between distinct phases as the level of homophily/heterophily is varied. As main findings, we unveil the emergence of an ‘overwhelming phase’ whose properties are analyzed in detail by a variety of analytical and computational methods presented in sections 4 and 5.

When the level of heterophily is non-existing or modest, the system is in an ‘ordinary phase’ similar to that characterizing the network with communities of equal size. Under intermediate to large heterophily, for sufficient asymmetry in the size of the communities, the agents of the majority group ‘overwhelm’ those of the minority by creating a

large number of cross-party links. We refer to this change of regime as the ‘overwhelming transition’, and to the regime itself as the ‘overwhelming phase’. In the overwhelming phase, the minority consists of agents having only cross-party links and large degrees following a broad distribution whose average can greatly exceed the preferred degree. By means of extensive Monte Carlo simulations and mean-field theories, we have determined the transition line separating the ordinary and overwhelming phases, and characterized in detail both regimes. In particular, we have studied the dependence on the level of homophily/heterophily and community size asymmetry of the number of cross-party links, fraction of agents with fewer links than the preferred degree, as well as the average degree in each community and the level of polarization in the network. In addition to these global quantities, we have also determined the total and JDDs of both communities.

We have found that the ordinary phase is characterized by features similar to those of the symmetric case [67]. The analysis of these follows from a direct two-community generalization of the mean-field approach used in the absence of group size asymmetry. The excellent agreement between simulation and analytical results has allowed us to show that in the ordinary regime, the network is essentially homogeneous, with total DD centred about the preferred degree and falling off exponentially (asymmetric Laplacian distribution), and with a broad distribution of cross-party links resulting in a ‘knife-edge’ JDD.

Remarkably, the overwhelming phase displays a number of surprising features: generally, the agents of the minority, all have a number of edges exceeding greatly the preferred degrees, and all of these are cross-party links. This results in a DD of the minority community that follows a broad Gaussian-like distribution. To characterize the latter, we have devised a nontrivial generalization of the ordinary-phase mean-field analysis which is found to be in good agreement with simulation data. Interestingly, the majority community in the overwhelming regime has essentially the same properties as in the ordinary phase: it forms a homogeneous network whose DD is centred about the preferred degree and that falls off exponentially. The transition from the ordinary to the overwhelming phase occurs at finite level of heterophily (when group sizes are asymmetric), and therefore differs from the fragmentation/fission, arising in other network models with homophily [27, 30, 32]. Such a transition, by which the network is split into disconnected communities, is also found in our model but only under extreme homophily.

It would be interesting to understand whether the existence of an overwhelming transition, the most distinctive features of our simple model, is robust against generalizations of our simple dynamic network model. As natural further avenues we could consider more than one preferred degree, or to allow agents to draw their preferred degree from a finite range. It would also be instructive to investigate other forms of update rules, e.g. like networks subject to heterophily and growing with preferential attachment [55]. An even more realistic, yet challenging, generalization would be to consider the co-evolutionary dynamics where network varies in response to changes of node states and the changes of those are coupled to updates of the network links. It would be quite relevant to investigate whether an overwhelming phase is a common feature of all these model extensions, and to what extent our analytical methods can

be generalized to tackle the latter. This endeavor, while challenging and likely to unveil even richer and more complex phenomenology, would allow us to shed further light on the important problem of better understanding the general features of dynamic network shaped by social interactions.

Acknowledgments

We are indebted and grateful to Andrew Mellor for substantial input and helpful discussions. The support of a joint PhD studentship of the Chinese Scholarship Council and University of Leeds to X L is gratefully acknowledged (Grant No. 201803170212). We are also grateful to the London Mathematical Society (Grant No. 41712) and Leeds School of Mathematics for their financial support, and R K P Z is thankful to the Leeds School of Mathematics for their hospitality at an early stage of this collaboration. This work was undertaken on ARC4, part of the High Performance Computing facilities at the University of Leeds, UK.

References

- [1] Asch S E 1955 Opinions and social pressure *Sci. Am.* **193** 31
- [2] Asch S E 1956 Studies of independence and conformity: I. A minority of one against a unanimous majority *Psychol. Monogr.* **70** 1
- [3] Schelling T C 1980 *The Strategy of Conflict* (Cambridge, MA: Harvard University Press)
- [4] Baronchelli A 2018 The emergence of consensus: a primer *R. Soc. Open Sci.* **5** 172189
- [5] Latané B 1981 The psychology of social impact *Am. Psychol.* **36** 343
- [6] Axelrod R 1997 The dissemination of culture: a model with local convergence and global polarization *J. Conflict Resolut.* **41** 203
- [7] McPherson M, Smith-Lovin L and Cook J M 2001 Birds of a feather: homophily in social networks *Annu. Rev. Sociol.* **27** 415
- [8] Yavaş M and Yücel G 2014 Impact of homophily on diffusion dynamics over social networks *Soc. Sci. Comput. Rev.* **32** 354
- [9] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591
- [10] Galam S 2012 *Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena* (New York: Springer)
- [11] Sen P and Chakrabarti B K 2013 *Sociophysics: An Introduction* (Oxford: Oxford University Press)
- [12] Mobilia M 2015 Nonlinear q-voter model with inflexible zealots *Phys. Rev. E* **92** 012803
- [13] Castellano C, Muñoz M A and Pastor-Satorras R 2009 Nonlinear q-voter model *Phys. Rev. E* **80** 041129
- [14] Mobilia M, Petersen A and Redner S 2007 On the role of zealotry in the voter model *J. Stat. Mech.* P08029
- [15] Albert R and Barabási A-L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47
- [16] Dorogovtsev S N and Mendes J F F 2003 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press)
- [17] Newman M 2010 *Networks* (Oxford: Oxford University Press)
- [18] Mellor A, Mobilia M and Zia R K P 2017 Heterogeneous out-of-equilibrium nonlinear q-voter model with zealotry *Phys. Rev. E* **95** 012104
- [19] Antal T, Redner S and Sood V 2006 Evolutionary dynamics on degree-heterogeneous graphs *Phys. Rev. Lett.* **96** 188104
- [20] Sood V, Antal T and Redner S 2008 Voter models on heterogeneous networks *Phys. Rev. E* **77** 041121
- [21] Baxter G J, Blythe R A and McKane A J 2008 Fixation and consensus times on a network: a unified approach *Phys. Rev. Lett.* **101** 258701
- [22] Blythe R A 2010 Ordering in voter models on networks: exact reduction to a single-coordinate diffusion *J. Phys. A: Math. Theor.* **43** 385003
- [23] Castellano C, Marsili M and Vespignani A 2000 Nonequilibrium phase transition in a model for social influence *Phys. Rev. Lett.* **85** 3536

- [24] Moretti P, Liu S, Castellano C and Pastor-Satorras R 2013 Mean-field analysis of the q-voter model on networks *J. Stat. Phys.* **151** 113
- [25] Szolnoki A, Perc M and Mobilia M 2014 Facilitators reveal the optimal interplay between information exchange and reciprocity *Phys. Rev. E* **89** 042802
- [26] Sabsovich D, Mobilia M and Assaf M 2017 Large fluctuations in anti-coordination games on scale-free graphs *J. Stat. Mech.* **053405**
- [27] Holme P and Newman M E J 2006 Nonequilibrium phase transition in the coevolution of networks and opinions *Phys. Rev. E* **74** 056108
- [28] Evans T S 2007 Exact solutions for network rewiring models *Eur. Phys. J. B* **56** 65
- [29] Vazquez F and Eguíluz V M 2008 Analytical solution of the voter model on uncorrelated networks *New J. Phys.* **10** 063011
- [30] Vazquez F, Eguíluz V M and Miguel M S 2008 Generic absorbing transition in coevolution dynamics *Phys. Rev. Lett.* **100** 108702
- [31] Lindquist J, Ma J, van den Driessche P and Willeboordse F H 2009 Network evolution by different rewiring schemes *Physica D* **238** 370
- [32] Durrett R, Gleeson J P, Lloyd A L, Mucha P J, Shi F, Sivakoff D, Socolar J E S and Varghese C 2012 Graph fission in an evolving voter model *Proc. Natl Acad. Sci. USA* **109** 3682
- [33] Henry A D, Pralat P and Zhang C-Q 2011 Emergence of segregation in evolving social networks *Proc. Natl Acad. Sci. USA* **108** 8605
- [34] Centola D 2011 An experimental study of homophily in the adoption of health behavior *Science* **334** 1269
- [35] Centola D and Macy M 2007 Complex contagions and the weakness of long ties *Am. J. Sociol.* **113** 702
- [36] Del Vicario M, Scala A, Caldarelli G, Stanley H E and Quattrocioni W 2017 Modeling confirmation bias and polarization *Sci. Rep.* **7** 40391
- [37] Centola D, González-Avella J C, Eguíluz V M and San Miguel M 2007 Homophily, cultural drift, and the co-evolution of cultural groups *J. Conflict Resolut.* **51** 905
- [38] Xie W-J, Li M-X, Jiang Z-Q, Tan Q-Z, Podobnik B, Zhou W-X and Stanley H E 2016 Skill complementarity enhances heterophily in collaboration networks *Sci. Rep.* **6** 18727
- [39] Ramazi P, Riehl J and Cao M 2018 Homophily, heterophily and the diversity of messages among decision-making individuals *R. Soc. Open Sci.* **5** 180027
- [40] Barranco O, Lozares C and Muntanyola-Saura D 2019 Heterophily in social groups formation: a social network analysis *Qual. Quant.* **53** 599–619
- [41] Yokomatsu M and Kotani H 2021 Knowledge sharing, heterophily, and social network dynamics *J. Math. Sociol.* **45** 111
- [42] Pariser E 2011 *The Filter Bubble: What the Internet is Hiding from You* (London: Penguin)
- [43] Iyengar S, Sood G and Lelkes Y 2012 Affect, not ideology: a social identity perspective on polarization *Public Opin. Q.* **76** 405
- [44] Barberá P, Jost J T, Nagler J, Tucker J A and Bonneau R 2015 Tweeting from left to right: is online political communication more than an echo chamber? *Psychol. Sci.* **26** 1531
- [45] Barberá P 2015 Birds of the same feather tweet together. Bayesian ideal point estimation using twitter data *Polit. Anal.* **23** 76
- [46] Bakshy E, Messing S and Adamic L A 2015 Exposure to ideologically diverse news and opinion on Facebook *Science* **348** 1130
- [47] Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley H E and Quattrocioni W 2016 The spreading of misinformation online *Proc. Natl Acad. Sci. USA* **113** 554
- [48] Wang X, Sirianni A D, Tang S, Zheng Z and Fu F 2020 Public discourse and social network echo chambers driven by socio-cognitive biases *Phys. Rev. X* **10** 041042
- [49] Shalizi C R and Thomas A C 2011 Homophily and contagion are generically confounded in observational social network studies *Socio. Methods Res.* **40** 211
- [50] McPherson J M and Smith-Lovin L 1987 Homophily in voluntary organizations: status distance and the composition of face-to-face groups *Am. Sociol. Rev.* **52** 370
- [51] Zeltzer D 2020 Gender homophily in referral networks: consequences for the medicare physician earnings gap *Am. Econ. J. Appl. Econ.* **12** 169
- [52] Gargiulo F and Gandica Y 2017 The role of homophily in the emergence of opinion controversies *J. Artif. Soc. Soc. Simul.* **20** 8
- [53] Boguná M, Pastor-Satorras R, Diaz-Guilera A and Arenas A 2004 Models of social networks based on social distance attachment *Phys. Rev. E* **70** 056122
- [54] Wong L H, Pattison P and Robins G 2006 A spatial model for social networks *Physica A* **360** 99

- [55] Karimi F, Génois M, Wagner C, Singer P and Strohmaier M 2018 Homophily influences ranking of minorities in social networks *Sci. Rep.* **8** 11077
- [56] Kimura D and Hayakawa Y 2008 Coevolutionary networks with homophily and heterophily *Phys. Rev. E* **78** 016103
- [57] Papadopoulos F, Kitsak M, Serrano M Á, Boguñá M and Krioukov D 2012 Popularity versus similarity in growing networks *Nature* **489** 537
- [58] Asikainen A, Iñiguez G, Ureña-Carrión J, Kaski K and Kivelä M 2020 Cumulative effects of triadic closure and homophily in social networks *Sci. Adv.* **6** eaax7310
- [59] Krapivsky P L and Redner S 2021 Divergence and consensus in majority rule *Phys. Rev. E* **103** L060301
- [60] Overgoor J, Benson A and Ugander J 2019 Choosing to grow a graph: modeling network formation as discrete choice *The World Wide Web Conf. (ACM)* pp 1409–20
- [61] Gorski P J, Bochenina K, Holyst J A and D’Souza R M 2020 Homophily based on few attributes can impede structural balance *Phys. Rev. Lett.* **125** 078302
- [62] Heider F 1958 *The Psychology of Interpersonal Relations* (Hove: Psychology Press)
- [63] Johnson N F, Xu C, Zhao Z, Ducheneaut N, Yee N, Tita G and Hui P M 2009 Human group formation in online guilds and offline gangs driven by a common team dynamic *Phys. Rev. E* **79** 066117
- [64] Liu W, Jolad S, Schmittmann B and Zia R K P 2013 Modeling interacting dynamic networks: I. Preferred degree networks and their characteristics *J. Stat. Mech.* P08001
- [65] Liu W, Schmittmann B and Zia R K P 2014 Modeling interacting dynamic networks: II. Systematic study of the statistical properties of cross-links between two networks with preferred degrees *J. Stat. Mech.* P05021
- [66] Bassler K E, Dhar D and Zia R K P 2015 Networks with preferred degree: a mini-review and some new results *J. Stat. Mech.* P07013
- [67] Li X, Mobilia M, Rucklidge A M and Zia R K P 2021 How does homophily shape the topology of a dynamic network? *Phys. Rev. E* **104** 044311
- [68] Krapivsky P L, Redner S and Ben-Naim E 2010 *A Kinetic View of Statistical Physics* (Cambridge: Cambridge University Press)
- [69] Liu W, Schmittmann B and Zia R K P 2012 Extraordinary variability and sharp transitions in a maximally frustrated dynamic network *Europhys. Lett.* **100** 66007
- [70] This can be derived by noting that $\mu_- \simeq \kappa$, so that $L_x \approx N_- \kappa \check{J}$. By estimating $\mu_+ \simeq L_x / N_+ = N_- \kappa \check{J} / N_+ = \mathcal{E} \kappa$, we have $\mathcal{E} = \mu_+ / \kappa \simeq \check{J} N_- / N_+$.
- [71] Prasetya H A and Murata T 2020 A model of opinion and propagation structure polarization in social media *Comput. Soc. Netw.* **7** 2
- [72] As a reminder, these quantities refer to steady state averages, and every expression involving ρ_σ has to be interpreted in the context of the mean-field approximation. Any referring to time dependence will be emphasized by, e.g. $\alpha_\pm(t)$.
- [73] We have assumed the fraction of connections in the network is small, meaning $\kappa \ll N$, so that an adder’s chance of finding an unconnected node is proportional to n_σ . Discrepancies between theory and data can be generally traced to the effects of finite size caused by this approximation.
- [74] In the example of figures 2 and 5 with $N = 1000$ and $\kappa = 60.5$, in the transition region for $|m| \ll 1$ and $J \lesssim J_c \approx -1$, we have found $\alpha_+ \in [0.024, 0.027]$ instead of being strictly equal to zero. Furthermore, in this example, $\mu_- \approx \kappa$ while $\mu_+ \gtrsim 70$, hence with $\mu_- - \mu_+ \gg 1$ rather than $\mu_- \approx \mu_+$.