Transmission errors and influence maximization in the voter model

Markus Brede, Valerio Restocchi, Sebastian Stein

University of Southampton, Department of Electronics and Computer Science, Southampton SO171BJ, UK

E-mail: `brede.markus@gmail.com`

Abstract. In this paper we analyze the effects of mistakes in opinion propagation in the voter model on strategic influence maximization. We provide numerical results and analytical arguments to show that generally two regimes exist for optimal opinion control: a regime of low transmission errors in which influence maximizers should focus on hub nodes and a large-error regime in which influence maximizers should focus on low-degree nodes. We also develop a degree-based mean-field theory and apply it to random networks with bimodal degree distribution, finding that analytical results for the dependence of regimes on parameters qualitatively agree with numerical results for scale-free networks. We generally find that the regime of optimal hub control is the larger, the more heterogeneous the social network and the smaller the more resources both available to the influencers.

1. Introduction

People often change their opinions in response to interactions with their peers or due to new knowledge from external sources of information. Such information contagion processes are relevant for many applications, ranging from political or government campaigns [1], marketing [2, 3], to the understanding of radicalization [4, 5]. In many of these contexts the question arises how an external agent with limited resources should best allocate its influence to maximally change opinion towards some target.

The latter question is traditionally addressed in the computer science and economics literature, mostly analyzing contagion processes related to the independent cascade model [6, 7, 8, 9]. Whilst the problem is NP-hard the model's simplicity and elegance have allowed for the development of fast computer algorithms with theoretical convergence guarantees that give good approximations for large scale social networks [6, 10, 11, 12, 13, 14, 15]. However, in the independent cascade model agents are assumed to be committed to an opinion, once they have adopted it. The model might thus be suitable for opinion spreading processes in which agents are strongly committed to a choice once adopted, like, e.g. buying expensive products or adopting certain life-style choices such as smoking.

The independent cascade model is not suitable when modeling opinion formation without strong commitment, e.g., when deciding about voting choices in elections or when agents can easily be swayed in their choices due to repeated influence from differing sources on social media. The dynamics of opinion change in the latter scenario are addressed in a rich interdisciplinary literature, cf. [16, 17] for reviews. However, questions of influence maximization on social networks have found relatively little attention for such dynamic models of opinion propagation. Exceptions are recent studies considering influence maximization for the case of continuous opinions [1] or for discrete opinions for Ising-like dynamics [18, 19, 20], a version of an AB model [21] and the voting dynamics [22, 23, 24, 25].

Here, because of its simplicity and prominence in the literature, we focus on the voting dynamics [26, 27], which has also been investigated in noisy settings [28]. The main effect of including noise in the voter model is that it prevents absorbing states, so no full consensus can be reached. Interestingly, varying trade-offs between the effects of noisy and pressure towards consensus formation from the copying dynamics, the noisy voter model exhibits a finite-size transition from a phase in which most of the agents spend most of the time close to one consensus state and then switch to the other to a phase in which two macroscopic subpopulations at different states exist [29, 30]. Recent work has shown that this finding essentially persists when interactions are given by complex networks [30].

Close to our work below, for the voting dynamics a large body of work has investigated effects of so-called zealotry on consensus formation [31, 32, 33]. Similar to what we call influencers below, zealots are agents that are less likely to change their opinions and thus they can exert substantial influence on the population. The effects of zealots has recently also been investigated in the noisy voter model for the mean-field scenario of all-to-all coupled populations [34]. However, literature on zealotry typically does not focus on the optimal allocation of a zealot's influence, and thus addresses a slightly different problem from the problem of influence maximization outlined below. However, recent work has also started to gain insights into optimal opinion control in the voting dynamics. Specifically, [22, 24, 25] have pointed out that heuristics targeting hub nodes perform very well on undirected heterogeneous networks, even though heuristics based on centrality metrics do not correspond to exactly optimal allocations for all types of networks [22]. This picture has been qualified by some of our recent work which has shown that targeting hub nodes is not necessarily optimal in all situations, e.g., not if agents have a large propensity to resist influence [35] or if time horizons of the influence maximizer are very short [36].

In the voting dynamics agents repeatedly adopt opinions of randomly selected neighbors. However, in real-world situations communication is often noisy, and we might not always correctly interpret a neighbor's choice or intention (or may be swayed by a variety of influences extrinsic to the system when making up our minds). Such a scenario of communication with frequently occurring misunderstandings might be particularly applicable in humanitarian situations [37]. Communication in such noisy

settings has previously been modeled in the so-called Kirman or noisy voter model [38, 28], but up to our best knowledge influence maximization has not yet been studied in the context of the noisy voter model.

Whilst strategies for optimal influence enhancement have been analyzed in various settings for the voting dynamics [22, 23, 24, 25], the effects of randomly occurring mistakes when adopting opinions on optimal allocation of external control have only found very limited attention. The only study in this domain is [20] who develop a gradient ascent algorithm based on mean-field solutions to study the role of noise in the Glauber dynamics for opinion formation. The authors of [20] find that targeting hub nodes is optimal in high noise scenarios while low-degree nodes should be targeted when noise is low.

Here we extend previous work on opinion control by investigating strategies for optimal influence allocation in the voting dynamics when opinion updating is subject to randomly occurring mistakes. Below, we extend the framework developed by [25] to this scenario and give guidelines how optimal allocation strategies have to be modified depending on the amount of uncertainty in communication.

Our approach to this problem is as follows. We first develop a numerical framework to estimate vote shares in the stochastic dynamics of the voter model via stationary solutions of the master equation. We then use optimization techniques to gain insights into optimal allocations, and develop an analytical understanding of the results via a degree-based mean-field theory. We also use these insights as the basis of testing heuristics which can easily scale to large-scale systems.

2. Model

Consider a set of N voters who each hold an opinion $s_i \in \{A, B\}, i = 1, ...N$. Agents are connected by a social network given by its adjacency matrix $G = \{a_{ij}\}_{i=1}^{N}$ which we assume to be symmetric and unweighted. Additionally, two influencers (or controllers) A (who holds opinion A) and B (who holds opinion B) are considered. Both can exert unidirectional influence on the N votes. This influencing is modeled through vectors $\vec{p}_A$ and $\vec{p}_B$ with $p_{A,i} = 1$ (or $p_{B,i} = 1$) if A (or B) exert influence on node i and $p_{A,i} = 0$ (or $p_{B,i} = 0$) otherwise. Note, that influence is thus modeled as a binary variable and –somewhat different to some approaches considering zealots [31]– the influencer is not subject to influence from the social system comprised of the N voters. Below we will always assume that A is active or strategic in the sense that A aims to allocate her influence with an aim to optimize her vote share in the steady state. In contrast, if present, B is considered passive in the sense of allocating her influence randomly.

Given some influence allocation $\vec{p}_A$ and $\vec{p}_B$ of the influencers we then proceed with the voter dynamics with transmission errors by iterating the following steps. (i) A randomly chosen focus agent selects a randomly chosen in-neighbor (which includes the controllers). (ii) The agent adopts the opinion of this in-neighbor with probability $1 - q$ and adopts the opposite opinion in the other case. The probability q thus models

'transmission errors', i.e. with probability q the message to follow an opinion is corrupted during transmission from the chosen in-neighbor to the focus agent. Steps (i) and (ii) are then iterated until a stochastic equilibrium is reached.

It is immediately obvious that for $q = 0$ the above model reduces to the traditional voter dynamics. However, it is worth commenting that our setup differs slightly from the traditional noisy voter model of [28] in which flip rates from states are augmented by a neighbour-independent probability additional to the ratios of neighbours of opposite states. The model as introduced above is meant to more explicitly model miscommunication as a source of noise, but could be mapped to the model of [28] with node-specific probabilities for state changes.

Note, that provided the social network is connected, for $q = 0$ the voting dynamics will always reach a consensus in finite networks [16]. For any $q > 0$ it is known that the noisy voter model is exponentially ergodic and an equilibrium state with certain fractions of voters adopting A and B votes will be reached. Importantly, other work on 'opinion control' in the voter model has assumed the presence of two influencing parties [25]. Without noise, such a choice is necessary because otherwise one opinion would become extinct. Here, the presence of transmission errors ensures the presence of an entropic force that always guarantees a mixed equilibrium and thus we can also consider settings of $\vec{p}_B = 0$.

The voting dynamics allows for analytical treatment [16]. Here we follow the approach of [25] and analyze it via mean-field rate equations for probability flows. Whilst the approach is a standard technique often used in this context, it is worthwhile bearing in mind that this involves an uncontrolled approximation that essentially ignores fluctuations, see, e.g. [39] for a more detailed discussion. Assume that x_i gives the probability of node i being in state A. We then have

$$\frac{dx_i}{dt} = (1 - x_i)\left(\frac{(1-q)(\sum_j a_{ij} x_j + p_{A,i})}{\sum_j a_{ij} + p_{A,i} + p_{B,i}} + \frac{q(\sum_j a_{ij}(1 - x_j) + p_{B,i})}{\sum_j a_{ij} + p_{A,i} + p_{B,i}}\right) -$$
$$- x_i\left(\frac{(1-q)(\sum_j a_{ij}(1 - x_j) + p_{B,i})}{\sum_j a_{ij} + p_{A,i} + p_{B,i}} + \frac{q(\sum_j a_{ij} x_j + p_{A,i})}{\sum_j a_{ij} + p_{A,i} + p_{B,i}}\right). \quad (1)$$

The terms in Eq. (1) correspond to probabilities of the focus agent chosen being of type B (top line) or A (bottom line) and then reflect probabilities of conversions according to the voting dynamics. For example, with the first term at the top line events in which a B is chosen as focus agent which will select an agent of type A as in-neighbor and correctly adopt its opinion state is modeled. With the second term on the top line an event in which a B chooses a B-agent to copy, but mistakes B's state and (mistakenly) adopts opinion A is described; analogous arguments lead to the terms at the bottom line. Here we are interested in stationary states of the opinion dynamics and hence set $dx_i/dt = 0$ for all i. After some simplification a linear equation for steady-state

probabilities $\vec{x}^*$ results

$$(\text{diag}(p_{A,i} + p_{B,i} + k_i) - (1 - 2q)G)\vec{x}^* = (1 - q)\vec{p}_A + q\vec{p}_B + q\vec{k}, \qquad (2)$$

where $k_i = \sum_j a_{ji}$ and $\text{diag}(p_{A,i} + p_{B,i} + k_i)$ gives a zero matrix with diagonal entries $p_{A,i} + p_{B,i} + k_i$. Note, that for $q = 0$ Eq. (2) reproduces Eq. (4) in [25]. The stationarity condition (2) gives a system of linear equation. Finding solutions to this system will allow us to optimize configurations of exerted influence below.

Before proceeding, we note that one can immediately read mean-field results for stationary vote shares for all-to-all connected networks from Eq. (2). Defining the average share of A opinions in the network by $X^* = 1/N \sum_i x_i^*$ we have

$$X^*(p_A, p_B) = \frac{(1 - q)p_A + qp_B + q}{p_A + p_B + 2q}, \qquad (3)$$

where $p_A = 1/N \sum_i p_{A,i}$ and $p_B = 1/N \sum_i p_{B,i}$ are the respective densities of influencers A and B. There is, of course, no point in influence maximization for an all-to-all connected network, but Eq. (3) confirms our intuition about the role of transmission noise. For $q = 0$ results reproduce results of [25], and for $q = 1/2$ we find $X^* = 1/2$ irrespective of p_A and p_B.

Similar to what was noted by Masuda [25] for the case without noise, we note that (2) is diagonally dominant for all q and thus, for shorter computation times for large networks, we can solve it by Jacobi iteration, starting, e.g., with $x_i^{(0)} = 1/2$, $i = 1, ..., N$ and then iterate

$$x_i^{(n+1)} = \frac{(1 - q)p_{A,i} + qp_{B,i} + qk_i + (1 - 2q)\sum_j a_{ji}x_i^{(n)}}{k_i + p_{A,i} + p_{B,i}} \qquad (4)$$

until convergence.

To obtain influence maximizing configurations we proceed as follows. First, undirected unweighted social networks are constructed using the configuration model [40] for power law degree distributions $P(k) \propto k^{-\alpha}$ where we typically consider $\alpha = 3$ in most numerical experiments below. More or less heterogeneous networks can be constructed by choosing smaller or larger values of the exponent α and adjusting proportionality constants to keep connectivities constant. In this procedure, multiple connections and self-connections are prevented and it is ensured that the resulting networks are connected. Next, for given influence resources $n_A = \sum_i p_{A,i}$ and $n_B = \sum_i p_{B,i}$ we allocate the influenced nodes by choosing n_A and n_B of the N nodes of the social network at random. Next, optimization of the influence of A from one randomly chosen influenced node i with $p_{A,i} = 1$ to another randomly chosen not yet influenced node j with $p_{A,j} = 0$ takes place if $X^*(p_{A,i} = 1, p_{A,j} = 0) < X^*(p_{A,i} = 0, p_{A,j} = 1)$. Influence reallocation according to stochastic hill-climbing is repeated until no further improvements have been found in the last $10N$ attempts.

In the above we have assumed that also the influence of controllers is subject to transmission errors. For comparison we have also considered a scenario where opinions
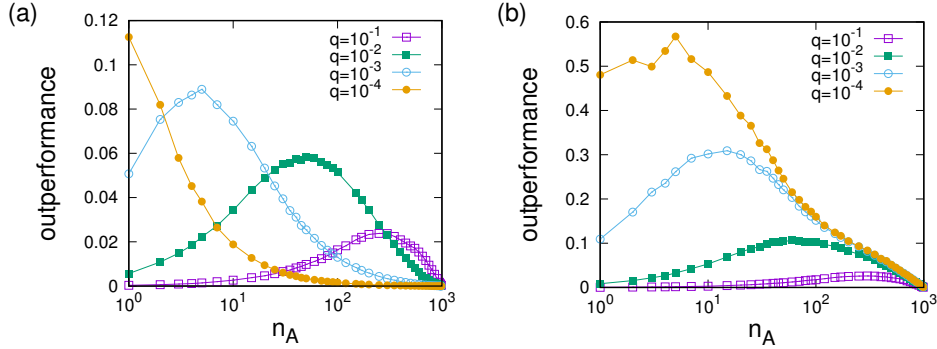
Figure 1. Dependence of vote share gains achievable by optimizing influence allocations $\vec{p}_A$ for influencer A on the influencer's total resource n_A for different amounts of transmission noise q. (a) Gains for A without a passive influencer B. (b) Gains for A against a passive influencer B with $n_B = 100$. Results are for scale-free networks with $\alpha = 3$ of size $N = 1000$ with average degree $\langle k \rangle = 3$. Data points represent averages over 50 network realizations and error bars are smaller or comparable to the size of symbols.

from the external influencer are always understood correctly. In that case Eq. (4) needs to be modified by omitting the pre-factor $1 - q$ to $p_{A,i}$ and the pre-factor q to $p_{B,i}$ on the right hand side. Results for both scenarios are essentially the same with small differences in numerical values for large n_A and n_B. To avoid unnecessary duplication we only present results for the first scenario.

3. Results

In this section we first present numerical results that illustrate a clear dependence of strategic influence allocation on heterogeneous complex networks on settings for transmission noise in subsection 3.1. Numerical results are complemented by analytical solutions for an illustrative network model in subsection 3.2, which clarifies the main observation of Sec. 3.1 and provides theoretical understanding.

3.1. Numerical Results

We first present results from numerical optimization in Fig. 1 which shows the dependence of influence gains achieved by optimization $X^*_{\text{opt}}/X^*_{\text{rand}} - 1$ on the budgets of the active and passive influencer for various amounts of transmission noise q. Several observations can be made. First, it is apparent that influence maximization can achieve larger gains in systems with low amounts of noise, whereas gains in very noisy systems are limited to single digit percentages. Gains also strongly depend on resource allocations to the competing influencer, i.e. acting against a well-resourced passive influencer a strategic influencer can generally improve outcomes by a far larger amount than when competing against a resource-poor opponent, when A's aim is essentially to
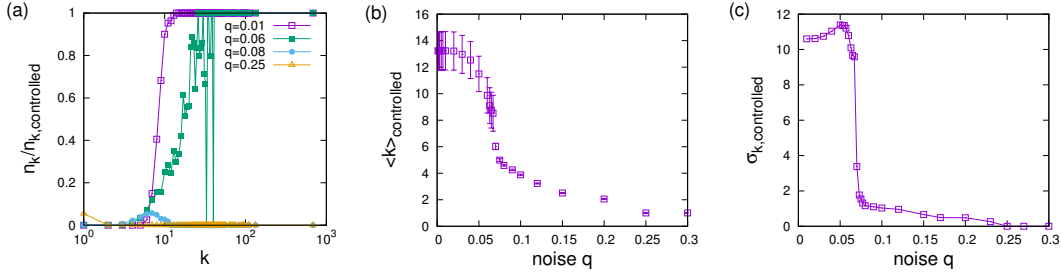
Figure 2. Strategic allocations of influence for different settings of the transmission error. (a) Dependence of likelihood of a node to be influenced by A on its degree k. (b) Dependence of average degree and (c) standard deviation of the distribution of influenced nodes in the optimal allocation $\langle k \rangle_{\text{controlled}}$ on transmission noise q. Parameters are $\alpha = 3$, $N = 1000$, $\langle k \rangle = 3$, $n_A = 10$, and $n_B = 0$ and data points represent averages over 50 runs. Note that for each setting of q optimizations have been performed for the same 50 networks.

dampen noise. Optimization gains also typically exhibit a maximum depending on A's resource endowment and the location of this maximum shows a clear dependence on transmission noise q. Generally we find that increasing q tends to shift the maximum towards larger n_A. Thus, in low noise settings strategic allocation matters most when resources of the active planner are small (e.g. $n_A = 1$ in Fig. 1a for $q = 10^{-4}$), but for the largest noise setting of $q = 1$ strategic allocation gives largest gains for large resource endowments (e.g. $n_A \approx 400$ in Fig. 1a for $q = 10^{-1}$). Results of these initial experiments thus clearly indicate major differences in influence maximization in the voter model with and without transmission noise.

To further understand the importance of transmission noise for strategic allocation we next constructed optimal allocations for different settings of q. We then quantify optimized allocations via the likelihoods of nodes to be influenced depending on their degree, i.e. the ratio $n_{\text{k,controlled}}/n_k$ (where n_k, $n_{\text{k,controlled}}$ are the number of nodes with degree k and the number of controlled nodes with degree k) and the distribution of influenced degrees. For the latter, two measures are of interest, e.g. the average controlled degree

$$k_{\text{controlled}} = 1/n_A \sum_i p_{A,i} k_i \qquad (5)$$

and the standard deviation of controlled degrees

$$\sigma^2_{\text{k,controlled}} = 1/n_A \sum_i p_{A,i} (k_i - k_{\text{controlled}})^2. \qquad (6)$$

The respective numerical experiments are illustrated in Fig. 2. Panel (a) shows the probability of nodes to be influenced depending on degree for low and high transmission error settings. For low error rate $q = 0.01$ the dependency is step like with all nodes with degree larger than $k \approx 20$ being subject to influence and all other nodes remaining uncontrolled. Thus clearly, for low transmission noise strategic influencers should almost
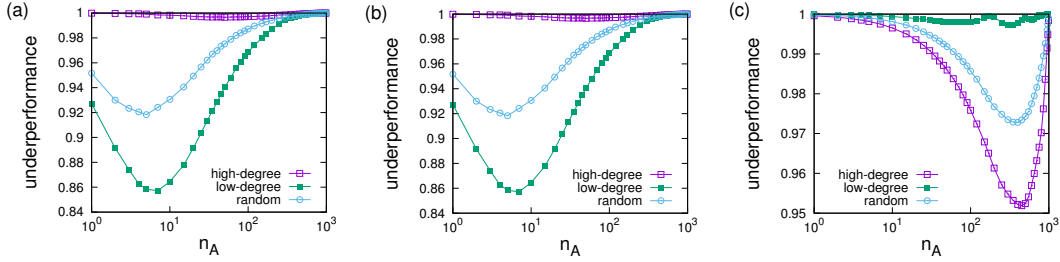
Figure 3. Comparison of performance of low and high degree heuristics relative to optimal allocations for settings with relatively low (left, $q = 0.01$), larger (middle, $q = 0.1$) and large (right, $q = 0.25$) transmission error probabilities. Parameters are $\alpha = 3$, $N = 1000$, $\langle k \rangle = 3$, $n_A = 10$, and $n_B = 0$ and data points represent averages over 50 runs.

exclusively target the highest degree nodes, which agrees with previous findings in the literature for the setting without noise [25]. As the error rate is increased this heuristic breaks down with a mixture of high and low degree nodes being targeted for $q = 0.06$, a focus on intermediate degree control for $q = 0.08$ and almost exclusive control of the lowest degree leaf nodes for $q = 0.25$. By showing the dependence of controlled degrees on q Fig. 2b reinforces this point. We observe a clear transition between a high-degree and low-degree control regime. In fact, exact transition points can easily be determined by plots of the standard deviation of the distribution of controlled degrees vs. the error probability q, cf. Fig. 2c, where we see a very sharp drop in the standard deviation of the distribution of controlled degrees when increasing q beyond the low-noise regime. As suggested in Fig. 2a the plot also confirms that the switch between exclusive hub control to leaf node control is not immediate. Instead, in between both phases we find a region of parameter space at approximately $0.07 < q < 0.25$ in which $\sigma^2_{k,\text{controlled}}$ is small, but does not vanish, corresponding to a situation where controlling nodes of intermediate and gradually declining degree maximizes vote shares.

The existence of separate low- and high-error regimes points out that heuristics based on high degree or centrality scores do not always fare best when considering opinion control in the voter model. In Fig. 3 we illustrate this point by comparing the relative under-performance $X^*_{\text{heuristic}}/X^*_{\text{opt}}$ of degree-based heuristics relative to optimal configurations. More specifically, we compare random allocations and low- and high-degree heuristics based on controlling the n_A highest or lowest degree nodes (and randomly choosing controlled nodes in case of degree ties). In agreement with [25], on undirected networks high-degree heuristics perform very well. However, note that (as also pointed out in [22]) exclusively controlling high-degree nodes does not always achieve optimal performance. For low error probability, controlling low-degree nodes in all settings proves inferior to controlling hub nodes. For high-error settings the exactly opposite result is observed, i.e. the low-degree heuristic leads to results very close to optimal, followed by random allocations and the high-degree protocol always performs worst. For the intermediate setting differences between the three protocols are subtle
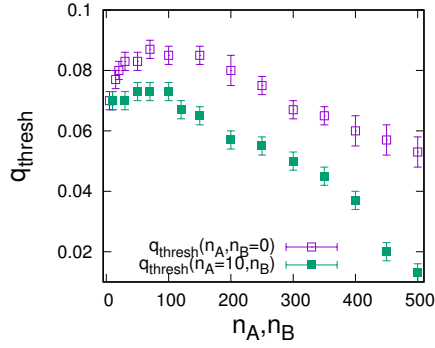
Figure 4. Dependence of the noise threshold that separates high- and low-degree control on agent resources. Open boxes give the dependence on the resource of the active influencer A when $n_B = 0$. Filled boxes give the dependence on the resource of the passive influencer when $n_A = 10$. Parameters are $\alpha = 3$, $N = 1000$, $\langle k \rangle = 3$, data points and errors estimated from plots of $\sigma_{k,\text{controlled}}$ vs. q.
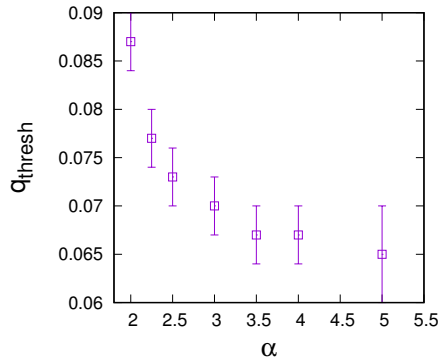
Figure 5. Dependence of the noise threshold on network heterogeneity measured by α. Parameters are $\alpha = 3$, $N = 1000$, $\langle k \rangle = 3$, $n_A = 10$, $n_B = 0$, and data points and errors estimated from plots of $\sigma_{k,\text{controlled}}$ vs. q.

and depend on resource endowments.

Our above result points out that simple degree-based heuristics can work reasonably well on heterogeneous networks, but require at least approximate knowledge about the error probability-control regime which applies in the system under investigation. To develop better intuition for this case we proceed by investigating the dependence of the low-high degree threshold q_{thresh} on resource allocations to the influencers and the network structure. As a good proxy for the threshold we use the level of noise at which the sharp drop in standard deviation of the distribution of controlled degrees is found (cf. Fig. 2c). We start by analyzing the dependence of the threshold on the resources n_A and n_B at the disposal of the active and the passive influencer, cf. Fig. 4. We consider two scenarios. In the first setting, we fix $n_B = 0$ and vary the resource n_A of A (open boxes in Fig. 4). In the second, we fix $n_A = 10$ and consider competition against an opponent B with varying resource n_B (filled boxes in Fig. 4). For both

settings a non-linear dependence of q_{thresh} and the respective resource endowments with a tendency for markedly lower thresholds the larger the resource budgets is observed.

In Fig. 5 we also investigate a setting in which we systematically vary network structure from very heterogeneous networks with scaling exponent $\alpha = 2$ up to more and more homogeneous networks as α is increased. Again, a clear dependence of the threshold on α is observed with values saturating at around $q_{\text{thresh}} \approx 0.065$ in the limit of regular networks. Thus, the more heterogeneous a network, the larger its high-degree control regime.

Both of our findings above give rough guidelines for selecting appropriate heuristics when some knowledge about the social network and opponent budgets is available. Roughly, the more resource available to the opponent and to the optimizer, the lower the noise threshold and the less appropriate hub-focused strategies are. Contrariwise, the more heterogeneous the social network, the larger the regime in which hub-control applies.

3.2. Degree-based mean-field theory

In the following we aim for an analytical understanding of the switch between optimal low- and high-degree control depending on the rate of transmission errors. For this purpose we approximate the voting dynamics with influence using a mean-field approach which will then allow us to find analytical expressions for optimal vote shares. We start with Eq. (2) which gives stationary vote shares of individual nodes and then group nodes according to degree k and whether they receive influence of strength one from A or not, which leaves us with $2k$ groups of nodes which we label $x_k^{(0)}$ (for nodes of degree k which are not influenced by A) and $x_k^{(1)}$ (for nodes which are influenced by A). For simplicity we don't explicitly model groups of nodes influenced by B or not, but assume that all nodes receive influence from B of average strength $p_{B,k} = n_B/N$ for all k. Then, introducing the shortcuts $\Delta_k^{(0)} = k + p_B$ and $\Delta_k^{(1)} = 1 + k + p_B$ and approximating $\sum_j a_{ij} x_j \approx k\langle x \rangle$, we obtain:

$$\Delta_k^{(0)} x_k^{(0)} = qk + (1 - 2q)k\langle x \rangle + qp_B, \tag{7}$$

$$\Delta_k^{(1)} x_k^{(1)} = 1 - q + qk + (1 - 2q)k\langle x \rangle + qp_B. \tag{8}$$

The mean field at the end of a randomly chosen link is

$$\langle x \rangle = \sum_k \frac{kh_k}{\langle k \rangle} \left(r_k x_k^{(1)} + (1 - r_k) x_k^{(0)} \right), \tag{9}$$

where h_k is the fraction of nodes with degree k and r_k gives the fraction of the nodes with degree k which are influenced by A.

Combining Eq.s (7) and (9) we obtain a self-consistency relation

$$\langle x \rangle = \sum_k \frac{kh_k}{\langle k \rangle} r_k \frac{1 - q + qk + qp_B + (1 - 2q)k\langle x \rangle}{\Delta_k^{(1)} k} + (1 - r_k) \frac{qk + qp_B + (1 - 2q)k\langle x \rangle}{\Delta_k^{(0)}} \tag{10}$$

which allows to solve for the mean-field

$$\langle x \rangle = \frac{\sum_k \frac{k h_k}{\langle k \rangle} \left(r_k \frac{1 - q + q k + q p_B}{\Delta_k^{(1)} k} + (1 - r_k) \frac{q k + q p_B}{\Delta_k^{(0)}} \right)}{1 - \sum_k \frac{k^2 h_k}{\langle k \rangle} (1 - 2q) \left(r_k / \Delta_k^{(1)} + (1 - r_k) / \Delta^{(0)} \right)}. \tag{11}$$

Equation (11) can now be used to find vote shares $x_k^{(0)}$ and $x_k^{(1)}$ for nodes with degree k via Eq. (7), which then gives an expression for the overall vote share

$$X = \sum_k h_k (r_k x_k^{(1)} + (1 - r_k) x_k^{(0)}). \tag{12}$$

We next seek to find influence allocations r_k that maximize X for given budget $n_A = N \sum_k h_k r_k$, given degree distribution h_k, and given rate of transmission errors q. Inspection of Eq. (12) with all substituted indirect dependencies shows that general expression for optimality quickly become extremely unwieldy. So, instead of proceeding with an analysis of general network structures, we restrict further analysis to a minimal network model, composed of a random network of equal fractions of hub nodes (with degree $k_1 > k_2$) and periphery nodes with degree k_2, so that the average degree is $\langle k \rangle = (k_1 + k_2)/2$. This simple network model is only a crude abstraction for heterogeneous complex networks, but allows to gain basic insights into optimal influence allocation to periphery and hub nodes.

For the above example, we can now evaluate Eq. (11) and obtain $\langle x \rangle = B/(1 - A)$ with

$$A = \frac{1 - 2q}{2\langle k \rangle} \left(r_{k_1} k_1^2 / \Delta_{k_1}^{(1)} + (1 - r_{k_1}) k_1^2 / \Delta_{k_1}^{(0)} + r_{k_2} k_2^2 / \Delta_{k_2}^{(1)} + (1 - r_{k_2}) k_2^2 / \Delta_{k_2}^{(0)} \right) \tag{13}$$

and

$$B = \frac{1}{2\langle k \rangle} \left(k_1 r_{k_1} \frac{1 - q + q k_1 + q p_B}{\Delta_{k_1}^{(1)}} + k_1 (1 - r_{k_1}) \frac{q k_1 + q p_B}{\Delta_{k_1}^{(0)}} \right) +$$
$$+ \frac{1}{2\langle k \rangle} \left(k_2 r_{k_2} \frac{1 - q + q k_2 + q p_B}{\Delta_{k_2}^{(1)}} + k_2 (1 - r_{k_2}) \frac{q k_2 + q p_B}{\Delta_{k_2}^{(0)}} \right). \tag{14}$$

Overall, we thus obtain

$$X = \frac{r_{k_1}}{2} \frac{1 - q + q k_1 + q p_B + (1 - 2q) k_1 \frac{B}{1 - A}}{k_1 + p_B + 1} + \frac{1 - r_{k_1}}{2} \frac{q k_1 + q p_B + (1 - 2q) \frac{B}{1 - A}}{k_1 + p_B} +$$
$$+ \frac{r_{k_2}}{2} \frac{1 - q + q k_2 + q p_B + (1 - 2q) k_2 \frac{B}{1 - A}}{k_2 + p_B + 1} + \frac{1 - r_{k_2}}{2} \frac{q k_2 + q p_B + (1 - 2q) \frac{B}{1 - A}}{k_2 + p_B}. \tag{15}$$

As we aim to divide a constant budget n_A between equal numbers of hub and periphery nodes we assume that a fraction $r_{k_1} = s n_A / N$ of the hub nodes and a fraction $r_{k_2} = (1 - s) n_A / N$ of the periphery nodes are influenced, so that varying $0 \leq s \leq 1$
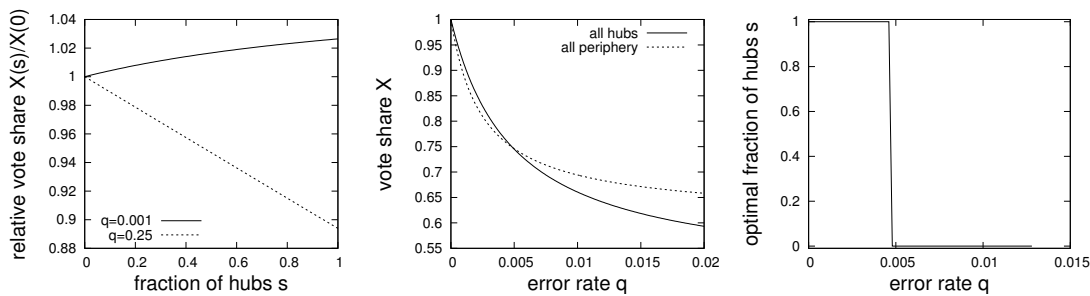
Figure 6. (a) Dependence of relative vote share $X(s)/X(s = 0)$ on the fraction of influenced hub nodes for a scenario of low transmission errors $q = 0.001$ and a scenario of high transmission errors $q = 0.25$. (b) Dependence of vote shares on transmission error rates q for an allocation in which all influence is assigned to hub nodes ($s = 1$), and an allocation of all influence to periphery nodes ($s = 0$). (c) Dependence of optimal fraction of influenced hub nodes on the error rate q for $n_A = 1$. Curves represent mean-field results based on Eq. (15) calculated for $k_1 = 100$, $k_2 = 1$, $n_A = N/2$, and $n_B = 0$.

allows us to interpolate between pure hub and pure periphery influence. Further, varying the fraction of overall influenced nodes n_A allows for a qualitative comparison with numerical results presented in Fig. 4 above. Substituting the corresponding terms into Eq. (15) gives an expression $X(k_1, k_2, n_A, n_B, s, q)$ that allows to compute vote shares as a function of all relevant parameters.

Using the expression just obtained for a concrete example network with $k_1 = 100$ and $k_2 = 1$ we illustrate the dependence of relative vote shares on the fraction of influenced hub nodes in panel (a) of Figure 6 for very low and large settings for transmission errors. Consistent with our earlier numerical results we clearly see that for large amounts of noise, the largest influence is obtained for $s = 0$, i.e. an exclusive allocation of all influence to periphery nodes. Contrariwise, for a low amount of errors, best vote shares can be obtained when most influence is allocated to hub nodes. Further results that show the dependence of vote shares obtainable for exclusive hub ($s = 1$) and exclusive periphery ($s = 0$) allocations on the amount of transmission noise are given in panel (b) of Fig. 6. Again, as observed in numerical optimization results in Sec. 3.1, we see a crossover from a low error regime in which hub allocation is optimal to a large error regime in which periphery allocation is best. Equation (15) also allows for a straightforward analytical derivation of an optimality condition and in panel (c) of Fig. 6 we also show the dependence of the optimal fraction of influenced hub nodes on q. In contrast to observations in the numerical optimization for scale-free networks above, results indicate an abrupt change from a regime of exclusive hub control to a regime of exclusive periphery control. The absence of a larger transition region in which a mix of high and low degree nodes is best likely is an artifact of the mean-field approximation which ignores differences in centrality between nodes.

Based on the observation that the transition between hub and periphery control is typically sharp, one can find good approximations for the dependence of the
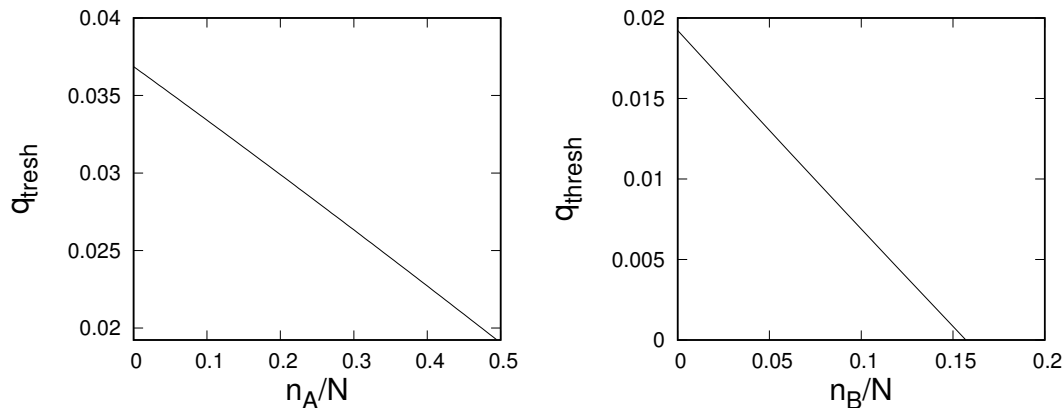
Figure 7. (a) Dependence of crossover point q_{thresh} between hub control to periphery control on the amount of resource n_A available to controller A for $n_B = 0$. (b) Dependence of crossover point q_{thresh} between hub control to periphery control on the amount of resource n_B available to controller B for $n_A = 1$. Settings are $k_1 = 20, k_2 = 5$.

transition point on various parameters by evaluating $X(k_1, k_2, n_A, n_B, s = 1, q_{\text{thresh}}) = X(k_1, k_2, n_A, n_B, s = 0, q_{\text{thresh}})$. Solutions of the resulting cubic equation are again involved and we only report illustrative results in Fig. 7. Panel (a) of Fig. 7 gives the dependence of q_{thresh} on the amount of resource available to controller A. As in the numerical optimization (cf. Fig. 4 – curve $q_{\text{thresh}}(n_A, n_B = 0)$) we see that thresholds are generally lower the larger n_A. In a similar vein, Fig. 7(b) reports the dependence on n_B for fixed $n_A = N/2$. Again, consistent with Fig. 4 in subsection 3.1 we find a decline in q_{thresh} with increasing resource availability for n_B.

4. Summary and Conclusions

In this paper we have analyzed the impact of noise in the form of randomly occurring misunderstandings in the voting dynamics on influence maximization. As one might expect, our results in the mean-field approximation and in numerical evaluation point out that networks become the more difficult to influence, the larger the probability of misunderstandings. Further, extensive numerical results for heterogeneous undirected networks combined with an analytical understanding obtained from a degree-based mean-field analysis point out that optimal allocation strategies will generally be different in low- and high noise-regimes which tend to be separated by a sharp transition. For low noise, we have shown that highest-degree heuristics give close to optimal allocations, while for high noise, lowest-degree heuristics give allocations within less than 0.5% off the optimum.

Interestingly, our findings in the dynamic setting of the voter model are in agreement with recent findings for optimal allocations in complex contagion dynamics [41]. In both situations it is found that nodes are the more difficult to control, the higher their degree.

In complex contagion this is the case in the beginning of cascades, when not enough activated neighbors are available to 'convince' them. In the voting dynamics with noise, nodes are exposed to the more potentially contradicting opinions the larger their degree, and thus the more difficult to influence the larger their degree.

References

[1] R. Hegselmann, S. König, S. Kurz, C. Niemann, and J. Rambau. Optimal opinion control: The campaign problem. *Journal of Artificial Societies and Social Simulation*, 18, 2015.

[2] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66. ACM, 2001.

[3] P. Domingos and M. Richardson. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth Intl. Conf. on Knowledge Discovery and Data Mining*, pages 61–70, 2002.

[4] M. Ramos, J. Shao, S. D. S. Reis, C. Anteneodo, J. S. Andrade, S. Havlin, and H. A. Makse. How does public opinion become extreme? *Sci. Rep.*, 5:10032, 2015.

[5] S. Galam and M. A. Javarone. Modelling radicalization phenomena in heterogeneous populations. *PLoS ONE*, 11:e0155407, 2016.

[6] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. *Proceedings of the Nineth International Conference on Knowledge discovery and Data Mining (KDD), Washington, DC, USA*, pages 137–146, 2003.

[7] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. *Internet and Network Economics, Lect. Notes Comput. Sci.*, 4858:306–311, 2007.

[8] S. Bharathi, D. Kempe, and M. Salek. Threshold models for competitive influence in social networks. *Proc. Workshop on Internet and Network Economics, WINE (2010)*, 4858:539–550, 2010.

[9] S. Goyal, H. Heidari, and M. Kearns. Competitive contagion in networks. *Games and Economic Behaviour*, 2014.

[10] E. Mossel and S. Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134. ACM, 2007.

[11] Y. Chen, W. Yuan and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 88–97. IEEE, 2010.

[12] W. Wang, C. Chen and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25:545, 2012.

[13] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 946–957. SIAM, 2014.

[14] Youze Tang, Xiaokui Xiao, and Yanchen Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 75–86. ACM, 2014.

[15] F. Morone and H. A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524:65–68, 2015.

[16] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

[17] A. Sirbu, V. Loreto, V. P. D. Servedio, and F. Tria. Opinion dynamics: Models, extensions and external effects. *Participatory Sensing, Opinions and Collective Awareness*, pages 363–401, 2016.

[18] S. Liu and S. Shakkottai. Influence maximization in social networks: An ising-model-based approach. *Proceedings of the Forty-Eighth Annual Allerton Conference*, pages 570–576, 2010.

[19] C. E. Laciana and S. L. Rovere. Ising-like agent-based technology diffusion model: Adoption patterns vs. seeding strategies. *Physica A*, 390:1139–1149, 2011.

[20] C. W. Lynn and D. D. Lee. Maximizing influence in an ising network: A mean-field optimal solution. *Proceedings of the 30th conference on neural information processing systems (NIPS 2016)*, pages 1–9, 2016.

[21] D. L. Arendt and L. M. Blaha. Opinions, influence, and zealotry: a computational study on stubbornness. *Comput. Math. Organ. Theory*, 21:184–209, 2015.

[22] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, 1:19, 2013.

[23] M. A. Javarone. Network strategies in election campaigns. *J. Stat. Mech.*, page P08013, 2014.

[24] C. J. Kuhlman, V. S. Anil Kumar, and S. S. Ravi. Controlling opinion propagation in online networks. *Computer Networks*, 57:2121–2132, 2013.

[25] N. Masuda. Opinion control in complex networks. *New Journal of Physics*, 17:033031, 2015.

[26] R. Holley and T. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3:643–663, 1975.

[27] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60:581–588, 1973.

[28] B. L. Granovsky and N. Madras. The noisy voter model. *Stoch. Process. Appl.*, 55:23–43, 1995.

[29] A. Carro, R. Toral, and M. San Miguel. Markets, herding and response to external information. *PLoS ONE*, 10:e0133287, 2016.

[30] A. Carro, R. Toral, and M. San Miguel. The noisy voter model on complex networks. *Sci. Rep.*, 6:24775, 2016.

[31] M. Mobilia. Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.*, 91:028701, 2003.

[32] M. Mobilia and I. T. Georgiev. Voting and catalytic processes with inhomogeneities. *Phys. Rev. E*, 71:046102, 2005.

[33] M. Mobilia, A. Petersen, and S. Redner. On the role of zealotry in the voter model. *J. Stat. Mech.*, page P08029, 2007.

[34] N. Khalil, M. San Miguel, and R. Toral. Zealots in the mean-field noisy voter model. *Phys. Rev. E*, 97:012310, 2018.

[35] M. Brede, V. Restocchi, and S. Stein. Resisting influence: How the strength of predispositions to resist control can change strategies for optimal opinion control in the voter model. *Frontiers in Robotics and AI*, 5:34, 2018.

[36] M. Brede, V. Restocchi, and S. Stein. Effects of time horizons on influence maximization in the voter dynamics. *Journal of Complex Networks*, page (in press), 2018.

[37] O. Okolloh. Ushahidi, or testimony: Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action*, 59:65–70, 2009.

[38] A. Kirman. Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 108:137–156, 1993.

[39] J. Yedidia. An idiosyncratic journey beyond mean field theory. *Advanced mean field methods: Theory and practice*, pages 21–36, 2001.

[40] M. E. J. Newman. *Networks: An introduction*. Oxford University Press, 2010.

[41] A. Alshamsi, F. Pinheiro, and C. Hidalgo. When to target hubs? strategic diffusion in complex networks. *arXiv:1705.00232v1*, 2017.