MPhys Project Theoretical Physics: Structure and Dynamics in Complex Networks

Opinion Dynamics on Networks with Community Structure

Franziska Klingner

supervised by Dr. N. Jones, Dr. M. Porter, Dr. D. Smith

*University of Oxford, UK**

(Dated: April 28, 2009)

The study of networks has concentrated on their complex (i.e. topologically non-trivial) structure on the one hand and dynamics on networks on the other hand. Only a little is known about the actual interplay of complex network structure and dynamics. We will investigate this for the case of opinion dynamics on social networks with community structure. Therefore, we work with two opinion models, the well-known voter model and the new "travelling competitors model" and examine the opinion spreading in the communities. Results from computer simulations (for both models) and analytical calculations (for the voter model) are presented and interpreted in the context of social integration.

1. INTRODUCTION

A *network* is an abstract representation of a system with a set of constituents and connections or interactions between them. Each constituent is identified with a *node* and each connection between two of those nodes is a *link*. Graphically nodes are denoted with dots and links with lines connecting those dots (see Fig. 1, 2 or 3). The generality of this network concept gives an impression why networks appear in so many areas, such as mathematics (here networks are called *graphs*), computer science, (natural) science and social science. See [1] for an introduction and [2, 3] for a review. Since this project is about opinion dynamics, we are dealing with social networks. Here the nodes represent individual persons and links a possible opinion exchange between them, for example because of a close friendship relation between the individuals. *Opinion*, the position of an individual towards a particular issue, is represented by assigning a value to each node. A review of possibilities for opinion values and dynamics can be found in [4]. In this paper, only models with just two possible opinion values or states b (black node) and w (white node) are considered. Not only does this simplify the analysis and numerics but also many decisions are in fact made between just two alternatives e.g. buying/ not buying a product or voting for Republicans/ Democrats [4]. Having a set of nodes with two possible states and interactions between them reminds of the Ising spin model for magnets. Indeed similar update rules as the aligning of an Ising spin with one of its neighbours have been used to model opinion dynamics [5]. A very popular one is the voter model, the details of which we will explain in section 2.2.1. This model has been applied on different network topologies [4], but mainly with the goal of examining the time it takes to reach a steady state, which is most often full *consensus* (all nodes have the same opinion value) [6–8]. This state corresponds to the ferromagnetic state of a magnet in physics or the extinction of a species in evolutionary biology. In social systems, for example people in a country voting for Republicans or Democrats, we rather observe coexisting different opinions and a change of the election result over time. Furthermore, people are often part of different groups, such as electoral districts, classes in society or ethnical groups [9]. Those groups are referred to as *communities* of the network, meaning that there are sets of nodes densely connected to each other, but only sparsely connected to other such sets [10]. In reality, the grouping of people can be related to their opinion. In the simulations, we can also see an influence of the community structure on the opinion dynamics for the voter model. This was studied in a summer project by James Wall [11]. He looked at the fractions of nodes with a particular opinion state in the different communities. We will call those fractions the *voting results* and denote it as β. Since it represents, what is mainly examined in opinion polls, this is a reasonable quantity to study. In our binary opinion models it is sufficient just to consider the voting result for one of the opinions (chosen to be b throughout this project) since it has to add up to one with the other voting result.

Equipped with this measure, we are able to perform quantitative studies of the combination of two key features of real opinion dynamical systems, the community structure and the behaviour far from the steady state. We do so by examining the spread of the voting results in the communities around the mean voting result in the network. Preliminary to this study we need to understand the considered networks (section 2.1), as well as the opinion models applied on them (section 2.2). The results for the different networks and opinion models are presented in turn in section 3. Finally, we draw a conclusion in section 4 and give an outlook to possible further work on the topic in section 5.

*Electronic address: franziska.klingner@hertford.ox.ac.uk ; Home university: Georg-August Universität Göttingen, Germany

2. MODELLING

2.1. Networks

This section will first explain the networks used in this study. Then we will have a look how networks can be represented uniformly and clarify what is meant with "community structure".

2.1.1. *Considered Networks*

1. To get an intuitive impression of the opinion dynamics we will apply them on a **two-dimensional lattice** (see Fig. 1) with periodic boundary conditions in both dimensions. The opinion configuration on such a network can easily be visualised as done later on in Fig. 4 and 8.

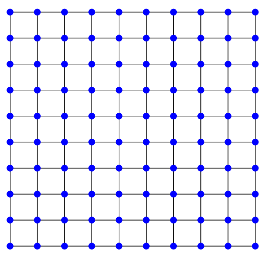

FIG. 1: Sketch of a two-dimensional lattice with edge length 10.

2. The next network should provide a basic community structure, but besides that be as simple as possible. That is why in this network each community has the same number of nodes (community size). Furthermore, each node has the same number of links attached to it. This number is called the *degree* of the node. To achieve a community structure, we use a part of the algorithm presented in [12]. We introduce a mixing parameter μ which determines the fraction of each node's links that connect this node to another community (*external links*). It follows that a fraction of $1-\mu$ of the links are *internal*, meaning those connect nodes to other nodes within the same community. The parameter μ will allow us to tune the strength of the community structure. Because of the homogeneous degree and community size distribution, we will call this the **homogeneous network** (see Fig. 2).

3. Moreover, we want to check the opinion dynamics on real data. Therefore, we use networks from the social networking website Facebook for various US universities (dated 2005). An example is shown in Fig. 3. A node represents a user and a link a "friendship" relation between two users, established through mutual confirmation of this status. The networks show a community structure which reflects the organisation at the universities [13]. On the one hand there are indeed votings taking place on these **Facebook networks**. On the other hand, we will treat those networks as surrogate for a non-virtual social networks. Because it is a great challenge to create artificial networks with a similar structure, those networks are considered a better approximation to non-virtual social networks. However, the exact quality is unknown.

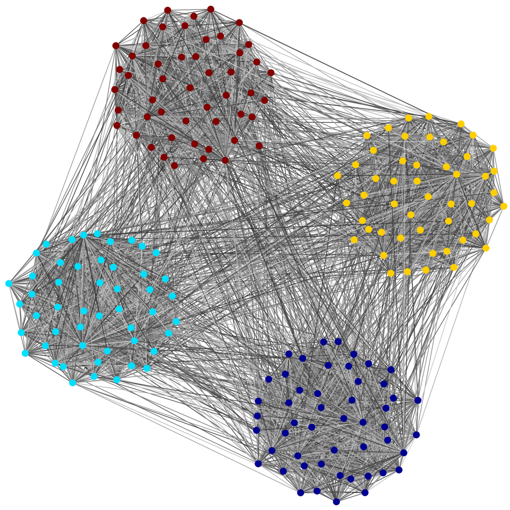

FIG. 2: Visualisation of the homogeneous network (200 nodes with degree 50) with community structure $\mu = 0.1$. The colours of the nodes represent their assignment to communities. The shades of grey of the links are random in order to improve the visibility as done in [10].

2.1.2. *Representation of Networks*

For analysis but not at least also for computations a network with N nodes can be represented by a $N \times N$ adjacency matrix $\mathbf{A}$ defined by $A_{ij} = 1$ if node i and j are connected by a link and 0 otherwise. In our networks we only allow an opinion exchange between different persons and the probability for an opinion exchange (link) should be symmetric. Therefore, the diagonal elements of $\mathbf{A}$ are all zero and the matrix is symmetric. Now we can state that $k_i = \sum_j A_{ij}$ is the degree of node i and $m = \frac{1}{2}\sum_{i,j} A_{ij}$ the total number of links in the network.

For a network with c communities, the assignment of the nodes can be written as an N-dimensional vector C (*index vector*) with integer entries between 1 and c labelling the communities. The opinion configuration is represented by an N-dimensional vector O with the entries 1 for b and 0 for w.

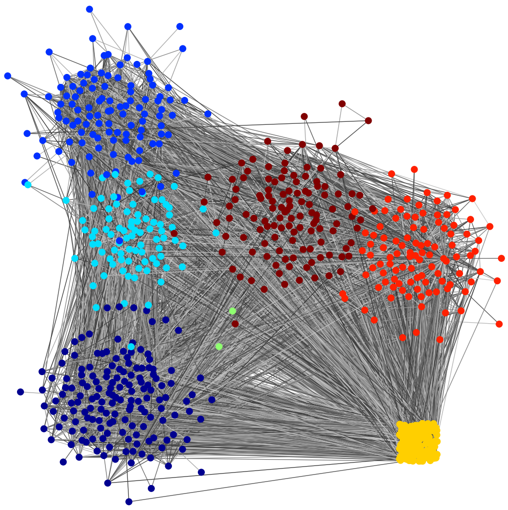

FIG. 3: Visualisation of the Facebook network of the Caltech University. Colours are assigned as in Fig. 2.

An example how to run opinion models (explained later in section 2.2) on a homogeneous network including the creation of the networks itself using MATLAB is given in Appendix A.

2.1.3. Community Structure

Since there is no strict definition of communities of a network, the applied method of finding them is crucial for whether we consider a network to have community structure at all. Hence it is important to clarify the method we used. Communities in the Facebook networks are found with the popular approach of modularity optimisation as proposed in [14]. The *modularity* Q measures the strength the community structure of a particular assignment $C = (c_1, c_2, \cdots, c_N)$ of nodes to communities. To do that, the quantity $I = \frac{1}{2\,m} \sum_{i,j} A_{ij}\ \delta(c_i, c_j)$ is considered which corresponds to the fraction of internal links in the network. It is compared to the expected fraction $E = \frac{1}{2\,m} \sum_{i,j} P_{ij}\ \delta(c_i, c_j)$ of those links in a randomized model of the network (null model) where P_{ij} is the probability to find a link between i and j therein. Now the modularity is defined as the difference of those two quantities.

$$Q = I - E = \frac{1}{2\,m} \sum_{i,j} [A_{ij} - P_{ij}]\ \delta(c_i, c_j) \qquad (1)$$

An often-used null model is the corresponding randomly-rewired network where the original index vector is kept and links are attached randomly such the degree distribution stays the same as for the original network.

As an example we can calculate the modularity of a homogeneous network with mixing parameter μ. Per construction the fraction of internal nodes is $I = 1 - \mu$. In the corresponding randomly-rewired network, the probability for a node of being connected to one the c communities is the same for all of them. Because each communities contains the same fraction of all nodes of the network, a fraction of $E = \frac{1}{c}$ links are expected to be internal. This results in

$$Q_{\text{hom}} = I - E = (1 - \mu) - \frac{1}{c} = \frac{c-1}{c} - \mu \qquad (2)$$

$$\text{with}\ \ \delta(x', x) = \begin{cases} 1 & \text{for } x' = x \\ 0 & \text{for } x' \neq x \end{cases}$$

Firstly, that shows that the modularity and the mixing parameter only differ by a sign and an additional constant what tells us that they are similar measures. Secondly, we can see that according to our intuition the modularity is maximized for a mixing parameter $\mu = 0$, hence for totally separated communities. Moreover we receive $Q = 0$ for $\mu = \frac{c-1}{c}$ where the network corresponds to a randomly-wired one. $Q \leq 0$, which implies $\mu \geq \frac{c-1}{c}$, corresponds to a network where the index vector does not reflect any community structure.

For the Facebook network the task of finding communities can be reformulated as finding an index vector which makes the community structure as distinct as possible, that means which maximizes the modularity. To perform that computationally, the Leading Eigenvector Method by Newman is used [15].

2.2. Opinion Dynamics

For our study only those models are of interest for which the dynamics far from the steady state show a dependence on the network's community structure

2.2.1. Voter Model

In this simple model the complex process of opinion formation in a social network is reduced to the fact that an individual might change its own opinion because it adopts the opinion of one of its friend with a different opinion. This is modelled by choosing two connected nodes of different opinion and copying the value from one to the other. In fact, there are three possibilities to perform this opinion update [6, 7]:

a) In the original *voter model* the node picked first adopts the opinion of one of its friends.

b) In the *reverse voter model* the node picked first spreads its opinion to one of its friends.

c) In the *unbiased voter model* or link-update model a link is picked first and it is chosen equally at random which one of the corresponding nodes adopts the opinion value of the other.

Opinion dynamics are created by repeatedly application of the update-rule. The difference in the dynamics implied by the three models just appears under inhomogeneous degree distributions [16]. Since that is not in our interest, we concentrate on the unbiased voter model where the dynamics is independent of a possible degree correlation of the two nodes attached to the picked link. Furthermore, we note that according to the definition above, only a certain class of links plays a role in performing an update, namely the links which connect nodes of different opinions, called *active links*. That reveals another advantage of the unbiased voter model: The probability of the change $b \to w$ happening in the network is the same as for $w \to b$ since every active link is attached to a white and a black node and it is randomly chosen which one changes its opinion to the other value. That allows us to write down the probabilities that the following changes happen in one update or time step:

$$P(b \to w) = P(w \to b) = \frac{1}{2} \,. \qquad (3)$$

2.2.2. Travelling Competitors Model

Instead of modelling opinion dynamics as an agreement process between friends, in this model the opinion is imposed on the individuals by a special person or object (e.g. the presidential candidate or the advertisement for a product). To spread the corresponding opinion value with respect to the network structure, we only allow a travelling of the person from friend to friend. That corresponds to a random walk on the network [16]. To ensure that the system stays away from consensus we take two of those walkers, one spreading opinion b and the other one opinion w, as they move independently through the network. As an example we could imagine Obama and McCain before the election, travelling to various locations in the different states and giving speeches to convince people to vote for them.
It is possible that both walkers move on and hit someone who already has the opinion they spread. We do not count that as an update. In all other cases the counter is incremented by one after both walkers took their step.

3. RESULTS

We will study the two opinion models on the three network types. First we look at the evolution of the opinion configuration on the two-dimensional lattice. After that we consider the homogeneous network and examine the influence of the community structure looking at the spread of the voting results. Finally we try to apply the models on the Facebook networks.

3.1. Voter Model on the Two-dimensional Lattice

We start examining the opinion dynamics by applying the voter model on the two-dimensional lattice with an edge length of 100 nodes. As initial opinion configuration we distribute the opinion values equally at random. This is shown in the top left panel of Fig. 4. The following two panels to the right show the opinion configuration after a certain number of updates. Therein the extensively studied coarsening process of regions with the same opinion value can be seen [17].

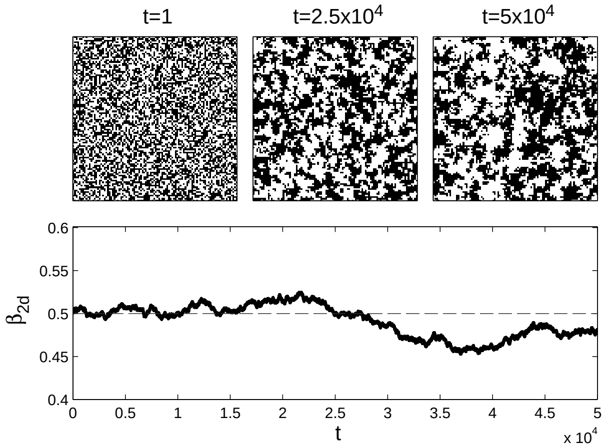

FIG. 4: Run of the voter model on a two-dimensional lattice with 100 × 100 nodes (top) Opinion configurations after $t = 1,\ 2.5 \cdot 10^4,\ 5 \cdot 10^4$ updates. (bottom) Fraction $\beta_{2\mathrm{d}}$ of black nodes in dependence of the number of updates t. The dashed line marks the initial value of $\beta_{2\mathrm{d}}$.

We quantify the dynamics by considering the voting result $\beta_{2\mathrm{d}}$, i.e. the fraction of black nodes in the network. Looking at possible time evolutions of $\beta_{2\mathrm{d}}$ (an example is shown in Fig. 4 in the bottom panel) we see that $\beta_{2\mathrm{d}}$ randomly fluctuates.

To determine the probability distribution of $\beta_{2\mathrm{d}}$, we have to reformulate the probabilities in Eq. 3 in terms of the change of $\beta_{2\mathrm{d}}$. The transition $w \to b$ means that we end up with one more black node than we had before. With N nodes in total in the network, that corresponds to an increase of $\beta_{2\mathrm{d}}$ by $\frac{1}{N}$. Equivalently $b \to w$ corresponds to $\beta_{2\mathrm{d}} \to \beta_{2\mathrm{d}} - \frac{1}{N}$. Equation 3 now reads $P(\beta_{2\mathrm{d}} + \frac{1}{N}|\beta_{2\mathrm{d}}) = P(\beta_{2\mathrm{d}} - \frac{1}{N}|\beta_{2\mathrm{d}}) = \frac{1}{2}$ and probability 0 for all other transitions. This can jointly be written using the Kronecker-δ as defined in Eq. 3

$$P({\beta_{2\mathrm{d}}}'|\beta_{2\mathrm{d}}) = \frac{1}{2}\,\delta({\beta_{2\mathrm{d}}}', \beta_{2\mathrm{d}} - \tfrac{1}{N}) + \frac{1}{2}\,\delta({\beta_{2\mathrm{d}}}', \beta_{2\mathrm{d}} - \tfrac{1}{N}). \qquad (4)$$

This is the master equation for a random walk on $\beta_{2\mathrm{d}}$ with step size $\frac{1}{N}$ [18]. Since we are dealing with networks with around $N = 1000$ nodes in this project, the step size is small $\frac{1}{N} \ll 1$ and Eq. 4 can be approximated by the corresponding Fokker-Planck equation for the probability distribution $p(\beta_{2d}, t)$, considering time as continuous [18] (see Appendix B).

$$\frac{\partial\, p(\beta_{2d}, t)}{\partial\, t} = \frac{1}{2}\ \frac{1}{N^2}\ \frac{\partial^2\, p(\beta_{2d}, t)}{\partial\, {\beta_{2d}}^2} \tag{5}$$

This is equivalent to a diffusion equation with diffusion constant $D = \frac{1}{N^2}$. The solution for the initial condition $\beta_{2d}(t = 0) = 0.5$ is known to be a Gaussian distribution.

$$p(\beta_{2d}, t) = \frac{1}{\sqrt{4\pi D t}}\ \exp\left(-\frac{(\beta_{2d} - \beta_{2d}(0))^2}{4Dt}\right) \tag{6}$$

That allows us to read off the expectation value and the variance.

$$\langle \beta_{2d} \rangle = \beta_{2d}(0) = 0.5 \quad \text{and} \quad {\sigma_{\beta_{2d}}}^2 = \frac{1}{N^2}\, t \tag{7}$$

We see that the standard deviation $\sigma_{\beta_{2d}}$ grows proportional to $\sqrt{t}$. Therefore, the system can reach consensus due to the high deviation from the mean for large t although the expectation value stays constant at 0.5. Since we want to study the behaviour away from consensus, we will choose t small enough that the probability of reaching consensus is kept small.

3.2. Voter Model on the Homogeneous Network

The homogeneous network cannot be visualised as demonstrative as the two-dimensional lattice, but again we can consider the voting result β. Since the homogeneous network has a community structure (at least for $\mu < \frac{c-1}{c}$ as defined in 2.1.3), we can furthermore look at the voting result in each of the c communities. We will denote the values as $\beta^* = \{\beta_1, \beta_2, \cdots, \beta_c\}$. Since we defined all community sizes to be equal in this network, the mean value $\overline{\beta^*}$ of the voting results in the communities is equal to the voting result for the whole network $\beta = \overline{\beta^*}$. Figure 5 shows an example how β and β^* evolve in time for two different values of μ.

We already analysed the curve representing the voting result in the whole network in section 3.1 (Eq. 6 and 7). Looking at Fig. 5, we note that in contrast to the curves for β, the curves for β^* seem to depend on μ. In the top panel of Fig. 5, we see that for a small value of μ, which corresponds to relatively sparsely connected communities, the curves for β^* spread widely around β. For higher values of μ however, we observe that all β^* curves stay quite close to the curve for β (Fig. 5 bottom). That fits to our intuition concerning real social networks: If groups of people are only sparsely connected, their opinion development is mainly independent of each other and the voting result in the different communities can differ a lot. Whereas, if the groups are highly connected, there is a lot of opinion exchange between them, resulting in just small differences in the voting result. Now our aim is to derive an analytical expression for the spread of the

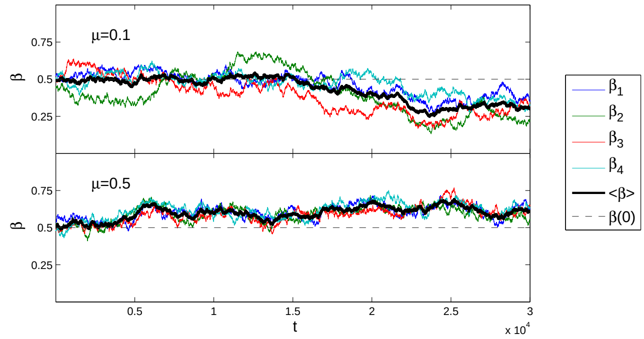

FIG. 5: Time evolution of the fraction of nodes with opinion b in the whole homogeneous network (β) and in the communities (β^*) for $\mu = 0.1$ (top) and $\mu = 0.5$ (bottom) and $N = 1000$, $c = 4$, $k = 30$.

voting results in the communities around the mean value, which means we have to find the standard deviation σ_{β_i}. Following the calculation of β_{2d} in section 3.1, we start by setting up the master equation for the voting result β_i in an arbitrary chosen community i. In the homogeneous network, all communities are equivalent so the statistics of β_i will not depend on the particular choice of the community i. To work with a variable which is normalized to mean value 0 we introduce $x = \beta_i - \beta$. This transformation does not change the standard deviation: $\sigma_x = \sigma_{\beta_i}$.

For determining the transition rates, we first have to notice that only with a probability of $\frac{1}{c}$ an update happens in the considered community i. Otherwise x does not change. This can be written as

$$P(x'|x) = \left(1 - \frac{1}{c}\right)\ \delta(x', x) + \frac{1}{c}\ P_i(x'|x) \tag{8}$$

We continue by deriving the transition probability $P_i(x'|x)$ for the change of x if the update happens in community i. Knowing that one node in i changes its opinion, we distinguish two possible reasons for that. Either the picking of an attached internal or an attached external link caused the opinion change. According to the definition of the homogeneous network the former has the probability $1 - \mu$ and the latter μ. If an internal link was picked, Eq. 3 is valid and β_i is randomly either increased or decreased by $\frac{1}{N/c} = \frac{c}{N}$ and so is x. Similar to Eq. 4 we can therefore write

$$\begin{aligned} P_i(x'|x) = &(1-\mu)\left(\frac{1}{2}\ \delta(x', x - \tfrac{c}{N}) + \frac{1}{2}\ \delta(x', x + \tfrac{c}{N})\right) \\ &+ \quad \mu \qquad\qquad P_{i,\mathrm{ext}}(x'|x)\ . \end{aligned} \tag{9}$$

In the last step we have to determine the transition probability $P_{i,\mathrm{ext}}$ when a change of x is caused by an external active link. We approximate the average fraction of black nodes in all other communities to be β.

The fraction of black nodes in community i however, is given by $\beta_i = \beta + x$. That yields the probability $l_{b,-} = (\beta + x)\,(1 - \beta)$ that a black node in i is connected to a white node in another community. The corresponding change $b \to w$ leads to a decrease of x. The probability for the opposite case of having a white node in i connected to a black one in another community has the probability $l_{w,+} = (1 - \beta - x)\,\beta$. Since we are dealing with probability one, we have to normalize $l_{b,-}$ and $l_{w,+}$ so that they add up to one. Approximating β by its expectation value $\beta \approx \langle\beta\rangle = 0.5$, we can write down the transition probabilities.

$$\begin{aligned} &P_{i,\text{ext}}(x'|x) = \qquad (10)\\ &\frac{l_{b,-}}{l_{b,-}+l_{w,+}}\,\delta(x', x - \tfrac{c}{N}) \;+\; \frac{l_{w,+}}{l_{b,-}+l_{w,+}}\,\delta(x', x + \tfrac{c}{N})\\ &\left(\frac{1}{2} - x\right)\delta(x', x + \tfrac{c}{N}) \;+\; \left(\frac{1}{2} + x\right)\delta(x', x - \tfrac{c}{N}) \end{aligned}$$

Putting together Eq. 8, 9 and 11, we finally get the complete master equation for the evolution of x.

$$\begin{aligned} P(x'|x) \;=\; &\frac{c-1}{c}\,\delta(x', x) \qquad (11)\\ +\;& \frac{1}{c}\,\delta(x', x + \tfrac{c}{N})\left[(1-\mu)\,\frac{1}{2} + \mu\,\left(\frac{1}{2} - x\right)\right]\\ +\;& \frac{1}{c}\,\delta(x', x - \tfrac{c}{N})\left[(1-\mu)\,\frac{1}{2} + \mu\,\left(\frac{1}{2} + x\right)\right]. \end{aligned}$$

As before, we deduce the Fokker-Planck equation (see appendix B).

$$\begin{aligned} &\frac{\partial p(x)}{\partial t} = -K\frac{\partial}{\partial x}\,x\,p(x) + \frac{1}{2}D\frac{\partial^2}{\partial x^2}\,p(x)\\ &\text{with} \quad D = \frac{c}{N^2} \quad \text{and} \quad K = \frac{2}{N}\,\mu \qquad (12) \end{aligned}$$

Equation 12 is the equation for an Ornstein-Uhlenbeck process [18]. It consists of a drift term linear in x with drift constant $K > 0$ resulting from having more active links attached to nodes with an opinion value which occurs more often than average. Furthermore we see a diffusion term with a diffusion constant D corresponding to a random walk as discovered in section 3.1. The solution $p(x, t)$ is quite complicated but the variance ${\sigma_x}^2 = {\sigma_{\beta_i}}^2$, we are interested in, takes a reasonable simple form [18].

$$\begin{aligned} {\sigma_{\beta_i}}^2 \;&=\; \frac{D}{2\,K}\,\left(1 - \mathrm{e}^{-2\,K\,t}\right) \qquad (13)\\ \Rightarrow \quad \sigma_{\beta_i} \;&=\; \sqrt{\frac{c}{4\,N\,\mu}\,\left(1 - \mathrm{e}^{-4\,\frac{\mu\,t}{N}}\right)} \qquad (14) \end{aligned}$$

The plot of the analytic results and the data from computer simulations (averaged over 20 runs) are shown in Fig. 6 where we can see that the values approximately match. The time values t are chosen such that they are large enough to minimize the influence of the initial condition and small enough that consensus is not reached.

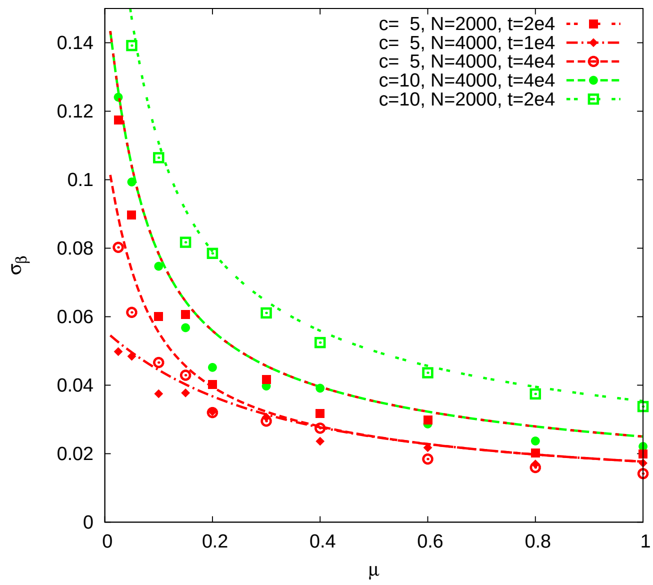

FIG. 6: Dependence of the standard deviation σ_β on the mixing parameter μ for the voter model ($k = 40$). Lines represent the graph of Eq. 14 and dots the results of the simulation.

Looking at Fig. 6 we first notice that our observation made in Fig. 5 that the spread increases for smaller mixing parameters proofs to be true. We already interpreted that as the more independent opinion development for groups of people with less opinion exchange between the groups. Thinking of the mixing parameter as a measure of social integration, we can now interpret the progression of the curves. They show a steep decrease for small values of μ whereas the slope decreases to zero as μ goes to 1. Therefore performing a change of $\Delta\mu$ in the range of sparsely connected communities, the effect in decreasing the opinion difference is relatively large. Whereas for communities which are already strongly connected (high μ value) the same $\Delta\mu$ leads to a small change in the opinion difference. Hence $\Delta\mu$ could be regarded as a measure for the amount of money investigated in integration and opinion exchange. Then, in order to get the maximum effect, integration programmes should concentrate on groups which are strongly separated.

Considering the dependence of the curves for a fixed N on the number of communities, we see that the spread becomes larger with increasing c for all values of μ. As an interpretation we could state that if a social network is split up in more communities than another one, we observe a bigger opinion spreading, even if the number of nodes and the mixing parameter are the same. Having a closer look at Eq. 14, we can note that for $\mu > \frac{N}{4\,t}$ (e.g. $\mu > 0.1$ for $N = 4000$, $t = 1 \cdot 10^4$) the exponential term becomes small. Thus, for that region the curve is mainly determined by the term $\sigma^2_{\beta,\text{approx}} = \frac{D}{2\,K} = \frac{c}{4\,N\,\mu}$. Firstly, this expression does not depend on time. Taking the limit of large times in Eq. 14 we see that $\sigma^2_{\beta,\mathrm{approx}}$ corresponds to the variance for the stationary solution. For the times we considered, the time dependence becomes important as μ becomes small as it can be seen in Fig. 6 looking at the two red curves for $t = 1 \cdot 10^4$ and $t = 4 \cdot 10^4$ which start to differ as μ becomes small. Secondly, $\sigma^2_{\beta,\mathrm{approx}}$ only depends on the fraction of nodes in each community $\frac{N}{c}$. That explains why we see the curves for $N = 4000$, $c = 10$ and $N = 2000$, $c = 5$ overlapping for high values of μ.

3.3. Voter Model on the Facebook Networks

In the last step of our study of the voter model, we will apply it on real social networks, represented by the facebook networks. Looking at the development of $\beta*$ (Fig. 7), we see similar curves as for the case of the homogeneous network (Fig. 5). That encourages us to draw predictions for those statistics from our results in section 3.2. To approximate the appropriate mixing parameter μ as good as possible, we first calculate the average value for each community and then we average over those values. Inserted into Eq. 14 together with the number of nodes of the network and the number of updates set to $t = N \cdot 10$, we can calculate the standard deviation of $\beta*$. We compare the result to the one obtained by averaging the standard deviation over 50 runs of the voter model on the network. The obtained values are shown in table I.

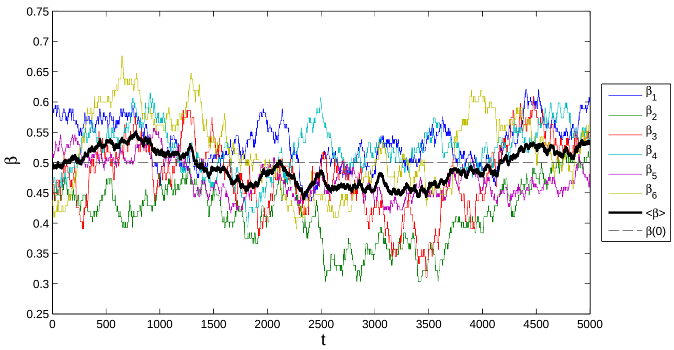

FIG. 7: Fraction of nodes with opinion b for running the voter model on the Facebook network "Caltech" in dependence of the number of updates t.

We see that we get reasonable predictions for some networks such as Reed or USFCA. The large relative deviation Δ_{rel} in the other cases could result from a large differences in the mixing parameters and community sizes in the network.

Network	N	c	$\langle k\rangle$	$\langle\mu\rangle$	$\sigma_{\beta,\mathrm{theo}}$	$\sigma_{\beta,\mathrm{sim}}$	Δ_{rel}
Caltech	769	6	43.3	0.42	0.057	0.068	0.19
Haverford	1446	7	82.4	0.56	0.072	0.046	0.36
Reed	962	3	39.1	0.31	0.053	0.05	0.06
Simmons	1518	3	43.5	0.25	0.054	0.045	0.17
Swarthmore	1659	11	73.6	0.60	0.248	0.053	0.79
USFCA	2682	6	48.7	0.40	0.037	0.037	0.01

TABLE I: Results for determining the standard deviation of the voting result $\beta*$ from theory (Eq. 14) and simulation.

3.4. Travelling Competitors Model on the Two-Dimensional Lattice

Now we switch to our second model. Again we start examining it on the two-dimensional lattice. In the top panel of Fig. 8 the opinion configuration for different times is shown. In contrast to the voter model, the travelling competitors model does not depend on the initial condition. That is why a possible initial condition of randomly assigned opinion values is shown in grey shades in Fig. 8. Running the model, both walkers create continuous regions of nodes with the opinion value they spread. We also saw those regions for the voter model (Fig. 4) and they have been growing with time. For the travelling competitors model, once those regions covered the lattice, their average size stays the same. Admittedly the walkers keep creating those regions in areas which previously were covered with the opinion of their competitor, but effectively only the position of the regions changes. Considering the fraction $\beta_{2\mathrm{d}}$ of black nodes plotted in the bottom panel of Fig. 8, we see fluctuation around $\beta_{2\mathrm{d}} = 0.5$ as for the voter model. But unlike there, these fluctuations now cannot lead to consensus since that would correspond to the extinction of one of the walkers.

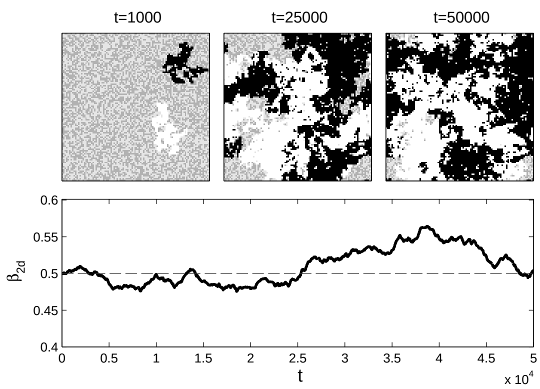

FIG. 8: Run of the travelling competitors model on a two-dimensional lattice (with periodic boundary conditions) with 100×100 nodes. (top) Opinion configuration after for different times. (bottom) Fraction of black nodes in dependence of time.

3.5. Travelling Competitors Model on the Homogeneous Network

We already saw that the community structure influenced the evolution of $\beta*$ for the voter model. Nevertheless the actual updates through picking an active link have always been the same. For the new model we might expect a change in the position of where successive updates happen in dependence of the network structure. Imagine for example a network with very sparsely connected communities. For the walker that means that there are only a few ways (links) which lead out of the current community. Thus, he stays longer in that community and spreads his opinion there meanwhile the other walker might be trapped in another community. Therefore we would expect the $\beta*$ curves to differ more from each other as μ decreases. We can actually see this process looking at the evolution of β as shown in Fig. 9. Note that the evolution after $t = 1 \cdot 10^4$ is plotted to ensure that the initial voting result has no influence any more. In the amber box the typical behaviour is well visible. The walker who spreads the opinion w in community 1, so that the blue curve for the fraction of nodes with opinion b decreases. The other walker spreads opinion b in community 2, so that the curve β_2 increases. In community 3 and 4 nothing happens, the voting result does not change.

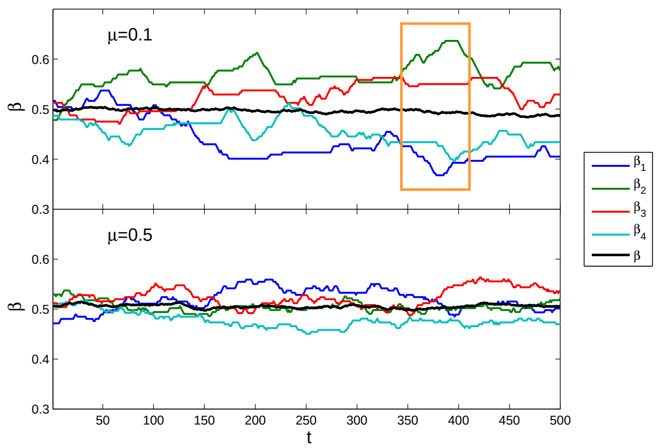

FIG. 9: Fraction of black nodes β in dependence of time for running the travelling competitors model on a homogeneous network with $\mu = 0.1$ (top) and $\mu = 0.5$ (bottom). See text for explanation of the behaviour (amber box).

Apart from that behaviour, we notice that the change of the spread depends on μ. To examine this we show the equivalent plot to Fig. 6 in Fig. 10.

Qualitatively we see the same behaviour as for the voter model. That means the interpretations we did in section 3.2 are valid for this model as well. Even quantitatively the values have approximately the same.

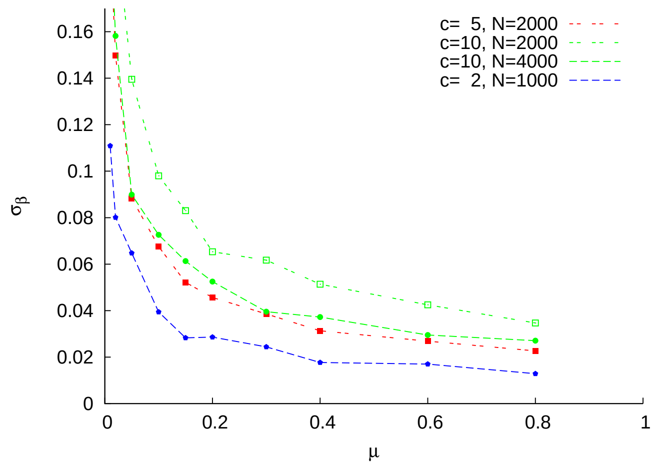

FIG. 10: Dependence of the standard deviation of $\beta*$ on the mixing parameter μ for the travelling competitors model ($k = 40$).

3.6. Travelling Competitors Model on the Facebook Networks

We recall that we found the community structure for the Facebook Networks using modularity optimization 2.1.3. Regarding the behaviour of the walkers we found in the previous section, we might think of another way of defining the community structure. We could say that a community is a set of nodes where a random walker has a higher probability to stay in the community as he moves along the links than when he would equally at random decide to stay or leave the community [19]. The latter corresponds to a probability of leaving the community of $P_l = 0.5$. For the homogeneous network this probability is $P_l = \mu$ since that is the fraction of external links attached to each node. Consequently, the resulting null model differs from the one used for finding the communities in the Facebook networks where we considered a network to have a community structure for $\mu < \frac{c-1}{c}$. Now the influence of the defining feature of "trapping the walker" is only visible for $\mu < 0.5$. Unfortunately, most of our real networks have mixing parameters around 0.5 (see I), so that an application of the travelling competitors model would not reveal the actual "positive" interplay of community structure and the spreading of the opinion values.

4. CONCLUSION

Applying two simple opinion models on complex networks we could quantify the influence of the community structure on the dynamics by considering the spread of the voting results in the various communities. Although both models had different update rules, we found a similar dependence on the community structure which fitted to our intuition about opinion dynamics in real networks.

We first derived the probability distribution and the standard deviation of the voting result for running the voter model on the two-dimensional lattice (Eq. 6). Performing a similar calculation for the homogeneous network, we then got an equation for the spread of the voting result in dependence of the mixing parameter μ (Eqn. 14). Furthermore we applied the voter model on our Facebook networks and have been able to provide good prediction for the resulting opinion spread (Tab. I) if the community sizes and mixing parameters in the Facebook networks are not too widely spread. For those cases eqn. 14 would also allow to determine the average mixing parameter for a network given the spread of the voting result, number of nodes and number of communities.

Applying the travelling competitors model on the two-dimensional lattice, we observed a similar behaviour of the voting result as for the voter model (Fig. 8). However, for the homogeneous network we found a much stronger dependence of the curves as for the voting results on the community structure. As a side effect of not being able to sensible apply the travelling competitors model on the real data we saw the importance of an appropriate community finding algorithm which fits to the considered dynamics.

5. OUTLOOK

Further studies would be necessary to determine the generality of the curves for σ_β that we found; we would have to observe the possible changes of this curve using different other opinion models. Those models could for example include random opinion flips or a continuous opinion variable. We also could investigate other measures of the influence of the community structure on the opinion dynamics, such as the mutual information or the correlation of the $\beta*$ curves [20]. Since those measures can include the time evolution of the data, a detailed study of the dependence on time and initial condition for both models would be possible. The results of such a study would allow a comparison of models to real data, for instance the time evolution of election results. As proposed in [11] gerrymandering phenomena could be studied therein by considering the dependence of the voting result on their assignment to communities of different sizes.

For the travelling competitors model, we could not express the evolution analytically so far. That would require further investigations.

Our formula for the spread of the voting result for the voter model could be improved by considering the degree distribution and the different community sizes.

Acknowledgements

I thank Nick Jones, Mason Porter and David Smith for their supervision, in particular for the freedom they gave me choosing the topic to work on. I acknowledge James Wall for the report on his summer project [11] which was an initiation for my ideas. Furthermore, thanks goes to Adam d'Angelo and Facebook, Inc. for providing network data which was suitable for parts of this project. Last but not least, I thank Peter Mucha et al. for providing the MATLAB code for community detection and visualisation on `http://netwiki.amath.unc.edu/`.

References

[1] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[2] M. E. J. Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics Review*, 45:167, 2003.

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports - Review Section Of Physics Letters*, 424:175–308, 2006.

[4] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. arXiv:0710.3256, 2007.

[5] S. Galam and S. Moscovici. Towards a theory of collective phenomena - consensus and attitude changes in groups. *European Journal Of Social Psychology*, 21(1):49–74, 1991.

[6] Claudio Castellano. Effect of network topology on the ordering dynamics of voter models. *American Institute of Physics Conference Proceedings*, 779:114, 2005.

[7] V. Sood, Tibor Antal, and S. Redner. Voter models on heterogeneous networks. *Physical Review E*, 77:041121, 2008.

[8] G. J. Baxter, R. A. Blythe, and A. J. McKane. Fixation and consensus times on a network: A unified approach. *Physical Review Letters*, 101(25):258701, 2008.

[9] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, June 2002.

[10] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. Communities in networks. arXiv:0902.3788, 2009.

[11] James F. Wall. Considerung opinion dynamics and community structure in complex networks: A view towards modelling elections and gerrymandering. Student report, presentation on `http://people.maths.ox.ac.uk/~porterm/temp/jimmy_final.ppt`, September 2008.

[12] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):046110, 2008.

[13] Amanda L. Traud, Eric D. Kelsic, Peter J. Mucha, and Mason A. Porter. Community structure in online collegiate social networks. arXiv:0809.0690, 2008.

[14] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.

[15] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3 Pt 2):036104, 2006.

[16] A. Barrat, M. Bartelemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge, 2002.

[17] E. Ben-Naim, L. Frachebourg, and P. L. Krapivsky. Coarsening and persistence in the voter model. *Physical Review E*, 53(4):3078–3087, Apr 1996.

[18] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, second edition, 1994.

[19] R. Lambiotte, J. C. Delvenne, and M. Barahona. Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770, 2008.

[20] Silviu Guiasu. *Information Theory with Applications*. McGraw-Hill, New York, 1977.

Appendix A

```
%%% MATLAB code for running the voter model
%%% on the homogeneous network

clear

%%% define parameters
N=1000;      % number of nodes
c=4;         % number of communities
up=10000;    % number of updates
k=40;        % degree of each node
mu=0.2;      % mixing parameter

%%% declare matrices for network representation
A=zeros(N,N);
  % adjacency matrix,
  % initially zero matrix of size NxN
C=randi(c,1,N);
  % index vector of length N,
  % filled randomly with label {1,2,...,c}
Op=randi(2,1,N)−1;
  % opinion vector of length N,
  % filled randomly with 0 (w) or 1 (b)

%%% declare auxiliary variables
data=zeros(up,c+1);
  % matrix of size (up x c+1) to save the voting results
  % for every time step
legs=ones(1,N)*k;
  % vector of length N with k in each entry,
  % represents desired homogeneous degree distribution,
  % used to count down free "legs" of a node
  % when establishing links later on
w=0;
  % counter for steps where nothing happend
  % when trying to create the network
```

```
%%% creating the homogeneous network

%   fill adjacency matrix

% continue establishing links
% as long as there are 2 legs left
% and it was not tried without success
% over 1000 times before
while sum(legs)>2 && w<1000
  % find the indices of all nodes
  % which still have free legs
  f1=find(legs>0);
  % pick a random one of those nodes
  i=f1(randi(size(f1,2)));

  % with a probability 1−mu
  % the node is connected to another node
  % in the same community
  if rand > mu
    % find indices of nodes in same community as i
    % unequal to i itself
    % with free legs
    % and not already connected to i
    f2=find(C==C(i) & legs>0 & 1:N≠i & A(i,:)==0);
    % stop if no such link exists
    if size(f2,2)==0; w=w+1; break, end;
    % else pick a random one of those nodes
    j=f2(randi(size(f2,2)));
  % else (probability mu) connect to a node
  % in another community
  else
    % find indices of nodes in another community as i
    % with free legs
    % and not already connected to i
    f2=find(C≠C(i) & legs>0 & A(i,:)==0);
    % stop if no such link exists
    if size(f2,2)==0; w=w+1; break, end;
    % else pick a random one of those nodes
    j=f2(randi(size(f2,2)));

  end

  % connect i and j with respect to A symmetric
  A(i,j)=1;
  A(j,i)=1;
  % reset counter for steps where nothing happend
  w=0;
  % count down the legs which are now occupied
  legs(i)=legs(i)−1;
  legs(j)=legs(j)−1;

end
```

```
%%% running the voter model

% loop through the update steps
for l=1:up,

  % print a warning if consensus is (nearly) reached
  % and exit the for−loop
  if sum(Op)==N−10 || sum(Op)==10
    display('consensus warning')
    break;
  end


  w=0;
  % find two connected nodes with different opinion
  while 1
    % choose a random node
    m=randi(N);
    % find all friends with different opinion
    f=find(A(m,:)==1 & Op≠Op(m));
    % break if such a node was found
    if size(f,2)≠0, w=0; break, end;
    w=w+1;
  end
  % choose a random one of those
  n=f(randi(size(f,2)));

  % perform opinion change
  % randomly from one node to the other one
  if rand>0.5
    Op(n)=Op(m);
  else
    Op(m)=Op(n);
  end

  % save current voting result in data
  % for each community
  for i=1:c
    f=find(C==i);
    data(l,i)=sum(Op(f))/size(f,2);
  end
  % and for the mean value
```

```
  data(l,c+1)=sum(Op)/N;
end
```

```
%%% running the travelling competitors model

% random start position of the walkers
r0=randi(N);
r1=randi(N);
% variable to remember whether an update happend
step=0;

% loop through updates
for l=1:up
  % reset step to zero
  % i.e. no update happend so far
  step=0;

  % walk with both walkers until
  % an update/ opinion change happens
  while 1
    % find friends
    f0=find(A(r0,:)==1);
    % walk to a random friend
    r0=f0(randi(size(f0,2)));
    % change its opinion
    % if it is not already the walker's opinion
    % and count update
    if v(r0)==1
      v(r0)=0;
      step=1;
    end

    % same for the other walker
    f1=find(A(r1,:)==1);
    r1=f1(randi(size(f1,2)));
    if v(r1)==0
      v(r1)=1;
      step=1;
    end

    % break up if a step was taken
    if step==1, break, end

  end
end
```

Appendix B

The Kramers-Moyal Expansion

The statistical processes we are dealing with is a Markov processes, i.e. the value of the random variable Y (the voting result in our case) at the next time step only depends on its value at the current time step. Furthermore, we can assume that our random variable has a continuous trajectory if the step size of a change of the variable is small. For those Markov processes it is generally possible to transform the equation for the transition probability $P(y', t'|y, t)$ into a Fokker-Planck equation for the probability distribution $p(y, t)$ [18]. The equation for this is given by

$$\frac{\partial\, p(y,t)}{\partial\, t} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{\partial}{\partial\, y}\right)^n \left[A_n(y)\; p(y,t)\right]$$

with moments

$$A_n(y,t) = \int \mathrm{d}y\; (y'-y)^n\; P(y', t+\mathrm{d}t|y,t).$$

In many cases only the first two terms are of interest:

$$\frac{\partial\, p(y,t)}{\partial\, t} = -\frac{\partial}{\partial\, y}\; A_1(y)\; p(y,t) + \frac{1}{2}\,\frac{\partial^2}{\partial\, y^2}\; A_2(y)\; p(y,t).$$

One-dimensional Random Walk

The transition probability for $t \to t + \Delta t$ is given by

$$P(y'|y) = \frac{1}{2}\,\delta(y'|y-l) + \frac{1}{2}\,\delta(y'|y+l) \tag{15}$$

where l is the step size of the random walk. That yields

$$\begin{aligned} A_1 &= 0 \\ A_2 &= l^2\;. \end{aligned}$$

Inserted in Eq. 5 we receive a diffusion equation because the first moment vanishes.

$$\frac{\partial\, p(y,t)}{\partial\, t} = \frac{1}{2}\; l^2\; \frac{\partial^2}{\partial\, y^2} p(y,t) \tag{16}$$

Ornstein-Uhlenbeck Process

Now we perform the same calculation for the master equation 12:

$$\begin{aligned} P(x'|x) &= \frac{c-1}{c}\;\delta(x',x) \\ &+ \frac{1}{c}\;\delta(x', x+\tfrac{c}{N})\left[(1-\mu)\,\frac{1}{2} + \mu\,\left(\frac{1}{2} - x\right)\right] \\ &+ \frac{1}{c}\;\delta(x', x-\tfrac{c}{N})\left[(1-\mu)\,\frac{1}{2} + \mu\,\left(\frac{1}{2} + x\right)\right]. \end{aligned}$$

Here we get

$$\begin{aligned} A_1 &= \frac{2\mu}{N}\;x \\ A_2 &= \frac{c}{N^2} \end{aligned}$$

and therefore

$$\frac{\partial p(x)}{\partial t} = -\frac{2\mu}{N}\frac{\partial}{\partial x}\; x\; p(x) + \;\frac{1}{2}\,\frac{c}{N^2}\,\frac{\partial^2}{\partial x^2}\; p(x)\;.$$

which corresponds to Eq. 12.