Triadic Closure Between Two Schools

AJ[a]

[a]Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK.

This manuscript was compiled on February 19, 2024

We generalize the work of Grindrod et al. [Internet Mathematics Vol. 8, No. 4: 402–423] on the dynamics of an edge-independent dynamic network. We model the evolution of friendships between two schools with triadic closure active and use a mean-field approach to approximate the macroscopic behaviour of the model in the long-term limit. Our numerical analysis shows that bistability generalizes to multi-stability with up to four stable fixed points and five unstable ones. Moreover, with the aid of three-dimensional vector plots, we visualize the macroscopic evolution of the model predicted by the mean field approximation, which agrees well with our simulations.

Evolving Networks | Triadic Closure | Mean Field Approximation | Markov Models

Networks, mathematical objects built of edges and vertices, provide a powerful class of models to study complex systems of interacting components. One can find a lot of information in a network [1, 2] and there exist many random graph models [2]. An important subfield of network science is that of dynamically evolving networks [3], as they are particularly relevant in real life scenarios. For example, epidemic networks modelling a disease are inherently dynamic and one would like to study the evolution rather than the final, static network [4]. Another example of evolving networks is one modelling human interactions. Such models can shine light on the connectivity structure of human behaviour. In [5], the idea of triadic closure is borrowed from social science [6] to model the evolution of interaction graphs in discrete time using Markov models. Grindrod et al. [5] use an edge-independent dynamic model using the framework introduced in [7]. They study mean-field approximations using the symmetries of the network to give macroscopic insights regarding the evolution of the connections. In particular, they treat the case where all the nodes are indistinguishable a priori and introduce the case where the vertex set is partitioned in two. In this paper, we seek to dive deeper in the latter model, thus modelling the evolution of the friendships between two schools for example. In the following section, we describe the mathematical setting and outline some results on the fully symmetric case. In the second section, we showcase how the model and its mean-field approximation generalize, and we attempt to find fixed points analytically. In the third section, we visualize the predictions given by the mean-field approximation, and we compare them to our numerical simulations. Finally, we give concluding remarks.

Preliminaries

Basic Definitions. In this paper, the setting for our analysis is identical to that of [5], which we include for completeness. We work with undirected graphs on n vertices with no self loops, represented by their adjacency matrices. An adjacency matrix $A = (a_{ij})_{i,j\in[n]}$ is an $n \times n$ symmetric matrix with a diagonal of zeros and off-diagonal elements in $\{0, 1\}$. We call the space of such matrices S_n and write $\mathbb{1} \in S_n$ for the adjacency matrix of the complete graph with n vertices. In general, the evolution of a graph can be determined by a stochastic rule generating the sequence $\{A_t\}_{t=1}^{T}$. We focus on a time homogeneous Markovian model, in the sense that conditioning on the full history of a process is the same as conditioning on the last step. Formally, given P_t the probability distribution on S_n at time t, we have

$$P_t(A_t|(A_1, \dots, A_{t-1})) = P_t(A_t|A_{t-1}).$$

By time homogeneous we mean $P_t = P$ for all t for some probability distribution P. Now let R_n be the set of symmetric $n \times n$ matrices with elements in $[0, 1]$ and define the expected value of A_t given A_{t-1} to be

$$\langle A_t|A_{t-1}\rangle \coloneqq \sum_{A_{t-1}\in S_n} A_t P(A_t|A_{t-1}) \in R_n$$

We say that a model is edge-independent if the presence of an edge does not affect the probability of another edge being present. Any matrix valued function $\mathcal{F} : S_n \to R_n$ generates an edge dependent evolving network model with expected value $\langle A_t|A_{t-1}\rangle$ at each $t = 2, \dots, T$. An intuitive form of $\mathcal{F}$ is taken from [7]

$$\mathcal{F}(A_{t-1}) = (\mathbb{1} - \tilde{\omega}(A_{t-1})) \circ A_{t-1} + \tilde{\alpha}(A_{t-1}) \circ (\mathbb{1} - A_{t-1}) \tag{1}$$

Significance Statement

The purpose of this paper is to study an evolving network modelling interactions between two schools. We assume friendships are created and are lost independently and that having friends in common with someone increases the chance of becoming friends. Furthermore, we apply a mean-field approach to study the macroscopic behaviour of the model and compare it to simulations.

where $\circ$ is the Hadamard product, $\tilde{\omega} : S_n \to R_n$ and $\tilde{\alpha} : S_n \to R_n$ are matrix valued functions representing the death rate and birth rate respectively.

Triadic Closure on Symmetric Graph. In [5], Grindrod et al. model triadic closure by letting $\tilde{\omega}(A_{t-1}) = \omega\mathbb{1}$ and $\tilde{\alpha}(A_{t-1}) = \delta\mathbb{1} + \epsilon A_{t-1}^2 \circ \mathbb{1}$ with $0 \leq \omega \leq 1$ and $\delta, \epsilon > 0$ such that $\delta + \epsilon(n-2) < 1$. Giovacchino et al. [8] argue that ϵ should scale like $\frac{1}{n}$ for births and deaths of friendship to be able to equilibrate. The intuition behind this formulation is that friendships die off independently with probability ω while new friendships are created with probability δ with an additional term depending on the number of friends in common. So, having friends in common increases the chances of becoming friends. In a mean field approach, one replaces an object of interest by its expected value, in this case A_t. Using the symmetries of the system, we obtain

$$\langle A_{t-1} \rangle = p_{t-1}\mathbb{1}.$$

Put that together with Equation (1) and $\mathcal{F}(A_t) = \langle A_t | A_{t-1} \rangle = p_t\mathbb{1}$, we get the mean field evolution

$$p_t = (1-\omega)p_{t-1} + (1-p_{t-1})(\delta + \epsilon(n-2)p_{t-1}^2) \tag{2}$$

Using the symmetry in the system, the dynamics on R_n are reduced to dynamics on $[0,1]$. In the limit where $\delta \downarrow 0$ and $\omega < \epsilon(n-2)/4$, this evolution has 3 fixed points, $\frac{\delta}{\delta+\omega} + O(\delta^2)$ and

$$\frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{\omega}{\epsilon(n-2)}} + O(\delta).$$

The middle point is found to be unstable, i.e. if we generate Erdös-Rényi random graphs with $p \approx p_{\mathrm{unstable}}$, we expect the edge density to flow towards either one of the stable fixed points. Simulations made by Grindrod et al. are consistent with this analysis, hence the "bistability" in the name of their paper. In the following sections, we seek to generalize the mean field approach outlined above to the case where the vertices are partitioned in two sets.

Triadic Closure Between Two Schools

Imagine instead of having indistinguishable vertices, we have two communities (schools) with n_1 and n_2 vertices (students) respectively. We simplify the edge evolution functions $\tilde{\omega}, \tilde{\alpha}$ in Equation (1) using three edge-independent friendship dynamics: the ones between the same school and the ones between the different schools. If triadic closure is active, we have three parameters for each of these dynamics. The question we seek to answer is how does the mean field approximation translate in this setting. Again, using [5], we have that the dynamics on R_n are reduced to dynamics on $[0,1]^3$. The three mean-field probabilities of interest are p, q and r, corresponding to the probability of friendship in the first school, the second school and

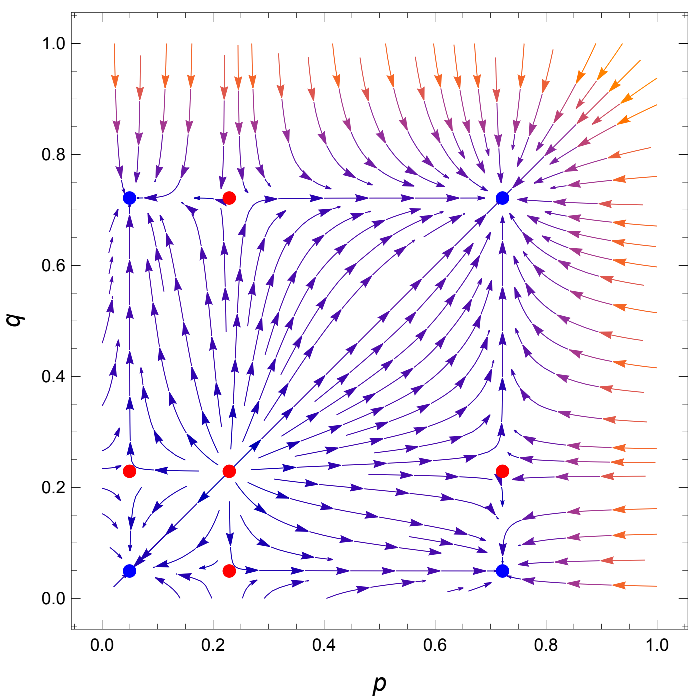

Fig. 1. Flow of the mean field evolution with no interactions between schools. The stable fixed points are shown in blue and the unstable ones in red. The parameters used are $\omega_p = \omega_q = 0.01, \delta_p = \delta_q = 0.0004, \omega_p = \omega_q = 0.0005, n_1 = n_2 = 100.$

between the two schools respectively. We use subscripts for the parameters accordingly. The mean field evolution is

$$\begin{cases} p_t &= (1-\omega_p)p_{t-1} + (1-p_{t-1})(\delta_p + \epsilon_p((n_1-2)p_{t-1}^2 + n_2 r_{t-1}^2)) =: f(p_{t-1}, r_{t-1}, \theta) \\ q_t &= (1-\omega_q)q_{t-1} + (1-q_{t-1})(\delta_q + \epsilon_q((n_2-2)q_{t-1}^2 + n_1 r_{t-1}^2)) =: g(q_{t-1}, r_{t-1}, \theta) \\ r_t &= (1-\omega_r)r_{t-1} + (1-r_{t-1})(\delta_r + \epsilon_r r_{t-1}((n_1-1)p_{t-1} + (n_2-1)q_{t-1})) =: h(p_{t-1}, q_{t-1}, r_{t-1}, \theta) \end{cases} \tag{3}$$

where θ is a vector of the parameters.

Independent Chains. Before getting into simulations, we would like to study this model analytically. Trying to find fixed points of this model is equivalent to solving a system of third degree polynomials with three variables and 11 independent parameters, which is futile without simplifications. We begin by forbidding interactions between schools, i.e. we set $\omega_r = 1, \delta_r = \epsilon_r = 0$, implying $r_t = 0$ for all t. We thus have two independent evolving chains which are bistable under the right conditions on the parameters. We show the flow diagram of $(f(p,0,\theta) - p, g(q,0,\theta) - q)^\top$ for two bistable copies of the chain taken from [5] in Figure 1. This vector field shown rather than $(f(p,0,\theta), g(q,0,\theta))^\top$ because the object of interest are fixed points and their stability. The flow in Figure 1 quantifies, for each point in $[0,1]^3$, the deviation from fixation given by the mean-field evolution. Therefore, following the arrows in such a flow diagram should lead us towards fixed points where $(f(p^*,0,\theta) - p^*, g(q^*,0,\theta) - q^*)^\top = (0,0)^\top$. The flow for two bistable copies, shown in Figure 1, tells us off the bat that the concept of bistability generalizes to multi-stability when we partition the vector set in two and look at the system as a whole. We note that the stability here is determined at each coordinate since the chains are independent. We expect more complicated flows in three dimensions when we turn on interactions between schools, but the concept of multi-stability to be present for some parameters.

Analytic Solutions. Ideally, one would like to find fixed points of the mean-field evolution and study their stability. But this is unreasonable without simplifications as there are 11 degrees of freedom in parameter space and up to at least 9 fixed points. When talking about fixed points, we obviously restrict ourselves to solutions in $[0,1]^3$. Now assume we have a fixed point $(p^*, q^*, r^*)^\top$ such that $f(p^*, r^*, \theta) = p^*, g(q^*, r^*, \theta) = q^*, h(p^*, q^*, r^*, \theta) = r^*$, its stability is determined using the connection with discrete dynamical system. We assume the parameters θ stay constant with time. Let $F : [0,1]^3 \to [0,1]^3 : (p,q,r)^\top \mapsto (f(p,r,\theta), g(q,r,\theta), h(p,q,r,\theta))^\top$, the vector field embedding the parameter evolution. The stability of $(p^*, q^*, r^*)^\top$ is determined by looking at the eigenvalues of the gradient of F at $(p^*, q^*, r^*)^\top$. If the maximal magnitude of these eigenvalues is strictly less than 1, then the fixed point is stable [9], which is a standard result. We tried many simplifications to reduce the degrees of freedom such as setting both schools to be identical, setting the interaction parameters as simple functions of the schools parameters. A mix of both approach reduces the number of degrees of freedom to 4 and still we were not able to find analytic solutions, even with the aid of Mathematica. Moreover, we found numerically that real solutions to the fixed points of F seen as defined on $\mathbb{R}^3$ need not even be in $[0,1]^3$ as was the case for Equation (2) where fixed points were bound to be in (0,1) (see [8, Theorem 3.1]). Therefore, even if one could guarantee a certain number of real solutions for the fixed points for some parameter values, there is the additional issue of the fixed point having to be in $[0,1]^3$. Having said this, it is helpful to visualize how parameters affect the number of fixed points and their stability. We see clearly in Figure 2 that

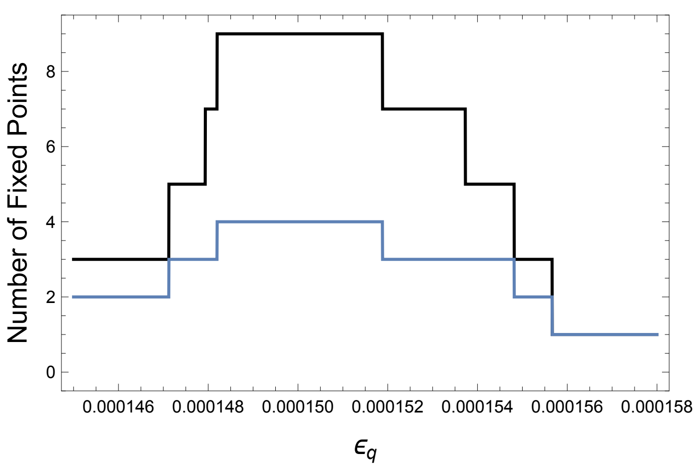

Fig. 2. Number of fixed points in $[0,1]^3$ as a function of ϵ_q with other parameters held fixed. Total number of fixed points are shown in black, and number of stable fixed points are shown in blue. Parameters used are $\omega_p = \omega_q = \omega_r = 0.003, \delta_p = \delta_q = 0.00015, \epsilon_p = 0.0001515, \delta_r = 0.00039, \epsilon_r = 0.0000065, n_1 = n_2 = 73$.

by slightly tweaking triadic closure within the second school, we greatly affected the number of fixed points. Moreover, we note that the stable fixed points that appear and disappear are not close to one another, suggesting sensitive dependence on parameters. This behaviour is by no means restricted to ϵ_q or this specific choice of parameters. Tweaking any parameter in this region of parameter space leads to seemingly counterintuitive fixed points solutions. This suggests that focusing on analytic solutions might prove uninstructive at best, and we therefore focus on numerical simulations. It is worth mentioning that other than showcasing the apparent complexity of analytical solutions, Figure 2 exposes an issue with the model that could prove problematic when using it on data. Indeed, to use this model, one would fit for the parameters as was done in [5].

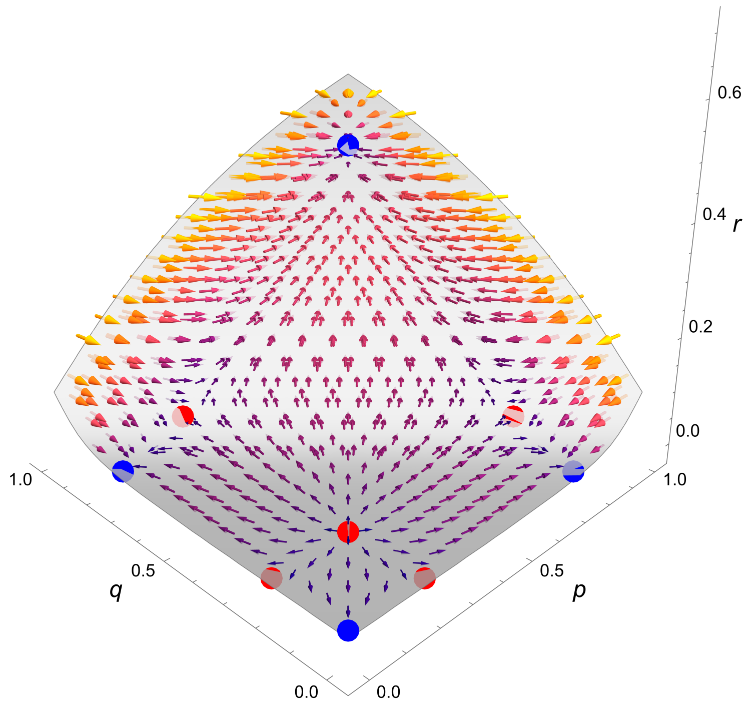

Fig. 3. Flow of mean-field evolution with interactions between schools on the surface where r is constant under the evolution. Stable fixed points are shown in blue and unstable ones in red. Vector sizes are scaled with respect to their magnitude. Parameters used are $\omega_p = \omega_q = \omega_r = 0.02, \delta_p = \delta_q = 0.0003, \delta_r = 0.0006, \epsilon_p = \epsilon_q = 0.015, \epsilon_r = 0.003, n_1 = n_2 = 73$.

But if we were to end up in a region of parameter space where this behaviour is prominent, errors on the fitted parameters due to statistical or systematic fluctuations in the data might lead to very different predicted dynamics. One should keep this in mind when fitting data and tweak parameters to see how fixed points solutions change.

Numerical Simulations and Predictions

Now that we have turned on interactions between schools, we would like to compare simulations to the mean-field evolution. We use edge density as in [5, 8] to have a sense of connectivity in the system as it evolves in time and for comparison with the mean field approximation of p, q and r. But in this setting, we need to define edge density in each block, i.e. for each school and between the schools. We define the block edge densities at time t as follows

$$\hat{p}_t := \frac{1}{\binom{n_1}{2}} \sum_{\substack{i,j=1 \\ i>j}}^{n_1} (A_t)_{ij}\,, \quad \hat{q}_t := \frac{1}{\binom{n_2}{2}} \sum_{\substack{i,j=n_1+1 \\ i>j}}^{n_1+n_2} (A_t)_{ij}\,, \quad \hat{r}_t := \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} (A_t)_{ij}$$

To simulate our model, we use the algorithm used in [5] starting with a random symmetric block matrix where entries are drawn from Bernoulli random variables with parameters p, q and r depending on the block. We paraphrase the procedure. At time $t-1$, we have adjacency matrix A_{t-1}. Each edge evolves as follows: if it is present, it disappears with probability $\omega_{p/q/r}$; if it doesn't exist, it is created with probability $\delta_{p/q/r} + \epsilon_{p,q,r}\#$(Number of common neighbours between the vertices), where $p/q/r$ is chosen depending on where the corresponding vertices lie. Having established the setup, we discuss some of our results. We choose some parameters so that we have four stable fixed points in the mean-field evolution. Figure 3 shows an example of the flow for such parameters projected on the surface $\left\{(p,q,r)^\top \in [0,1]^3 : h(p,q,r,\theta) = r\right\}$. Without projection, it is difficult to visualize the complex flows in three dimension. It is only natural to choose a surface containing all the fixed points as we would like to see how the flow from unstable fixed points connects to stable fixed points. For non-trivial parameters, we find the surface with h constant more visually adequate than the surfaces with f or g constant. This is true because of the explicit forms of f, g and h. Indeed, we see in Equation (3) that h has at most a quadratic dependency on a variable when keeping the others fixed, while f and g have cubic dependencies on p and q respectively. This leads to a surface with less curve, thus more tractable, yet containing all the interesting points. As seen in Figure 3, we observe 3 stable fixed points with r^* close to 0, i.e. low socialization between schools and one with $r^* \approx 0.44$. Moreover, the unstable fixed point at approximately $(0.23, 0.23, 0.006)$ has flows around it going towards the four stable fixed points via direct routes and via other unstable fixed points. The steepness of the surface starting past the $p = q$ line in the area where $r \approx 0$ suggests that for these parameter values, the two schools cannot simultaneously have high inner socialization without being interconnected. Now, when starting our simulations from random block matrices where $(p,q,r)^\top$ were not fixed points of Equation (3) (or close to one), the mean field evolutions approximated the block edge densities well, with the convergence to the same fixed point every time, as predicted by the flow. When starting from a stable fixed point, we found that the edge densities stayed constant which

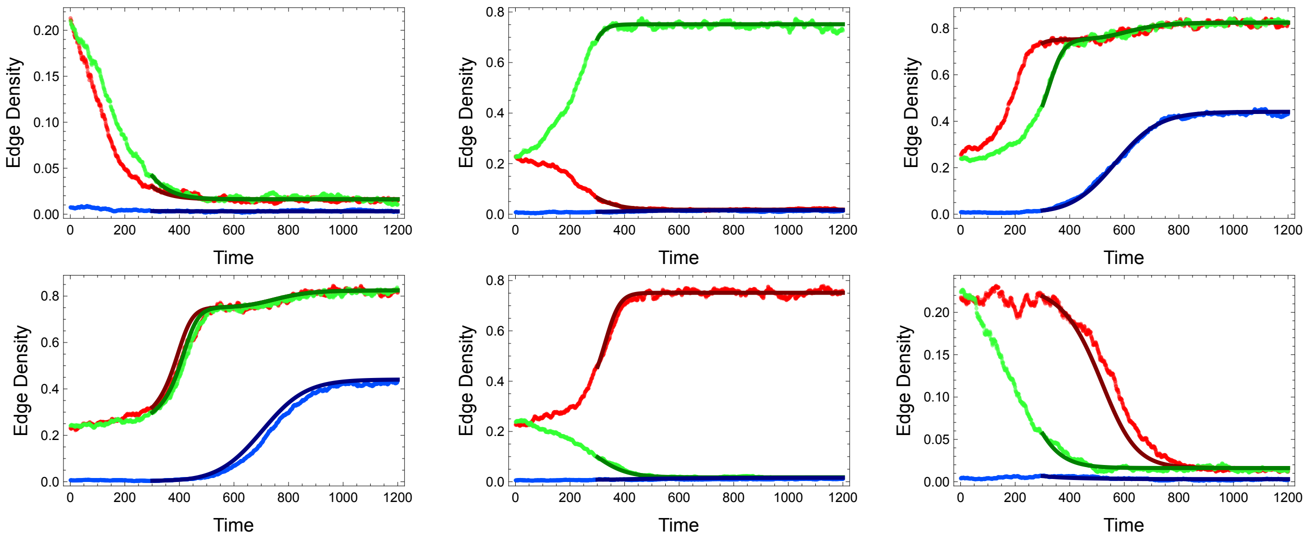

Fig. 4. Simulations (1200 time steps) of the evolving block model starting from a random block matrix with parameters (0.23,0.23,0.006), which is the central unstable fixed point from Figure 3. The mean-field approximation is turned on from time step 300 onwards. Block edge densities $\hat{p}_t$, $\hat{q}_t$ and $\hat{r}_t$ are shown in red, green and blue respectively.

is clearly expected from a stable fixed point. Multi-stability comes into play when the dynamics start from an unstable fixed point, which is unsurprising given that this is what was observed in the fully symmetric case [5]. Again, looking at the flow of the vector field, one can predict the possible avenues the system can take. But this is not to say that we can predict the long term behaviour at every run. Indeed, that's where the stochasticity is the most pronounced as can be seen in Figure 4. We clearly observe behaviours predicted by the mean-field approximation, although there's no way of knowing a priori which route the model is going to take. The evolutions toward the symmetric and asymmetric (in p and q) stable fixed points are shown in the first and second columns respectively. In the third column, we observe the evolutions towards stable fixed points via alternative routes, namely near unstable fixed points. Figure 4 shows that, after a sufficient amount of time to escape the highly stochastic region, the mean-field approximation models the macroscopic behaviour really well. Moreover, we every route from unstable fixed point to stable fixed point in Figure 3 is observed in our simulations. It is safe to say that the mean-field approximation is an effective way to predict possible macroscopic behaviour for this model.

Some Remarks. We note that in our simulations, we couldn't find parameters such that there were more than four stable fixed points and nine total fixed points, suggesting that the example presented in this section covers comprehensively the interesting behaviour of the model. Moreover, by changing the parameters so that both schools have different dynamics, we find for some instances four stable fixed points in which each coordinate has a unique value. This implies that the concept of bistability indeed generalizes to multi-stability, in the sense that a school can have more than two stable edge densities. As an aside, the stochasticity of the model makes it possible for a system near a stable fixed point to suddenly deviate, but we never encountered such an instance. This is to be expected; Giovacchino et al. [8] estimate that the mean exit time in a similar model from a stable state to another is approximately exponential in the number of vertices. A final remark is that just like in the fully symmetric case, a lack of stimulus in the early stages of the dynamics might lead to disappearance of socialization virtually forever (given that we have high mean exit time in principle).

Discussion

In this paper, we analysed a dynamically evolving network modelling the interactions between schools with triadic closure active. We studied the extension of the mean-field approximation to the model. We found that unlike the fully symmetric case, finding analytic solutions to the fixed point system of equations was challenging and highly sensitive to parameters. However, numerical exploration of the fixed point solutions show that the concept of bistability generalizes to multi-stability. We conjecture that for any set of parameters in this model, there are at most four stable fixed points. By looking at the fixed point flows from the mean-field approximation, we were able to visualize how the macroscopic behaviour of the model can evolve, which showed to be in close agreement with our simulations. This class of model using mean-field approximation shows great promise for modelling connectivity of evolving networks, and its application to real data could provide additional insights that were missed in this report.

ACKNOWLEDGMENTS. We would like to thank Stephen Wolfram and his team for building a great software that keeps on delivering.

1. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10):P10008, oct 2008. .
2. Mark Newman. Measures and metrics. In *Networks*, chapter 7,11. Oxford University Press, 07 2018. ISBN 9780198805090. .
3. Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, oct 2012. .

4. Jing Cui, Yi-Qing Zhang, and Xiang Li. On the clustering coefficients of temporal networks and epidemic dynamics. In *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2299–2302, 2013. .
5. Peter Grindrod, Desmond Higham, and Mark Parsons. Bistability through triadic closure. *Internet Mathematics*, 8, 12 2012. .
6. Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360 – 1380, 1973.
7. Peter Grindrod and Desmond J. Higham. Evolving graphs: dynamical models, inverse problems and propagation. *Proceedings: Mathematical, Physical and Engineering Sciences*, 466(2115):753–770, 2010. ISSN 13645021.
8. Stefano Di Giovacchino, Desmond J. Higham, and Konstantinos C. Zygalakis. A hierarchy of network models giving bistability under triadic closure. *CoRR*, abs/2111.05715, 2021. URL https://arxiv.org/abs/2111.05715.
9. Hiroki Sayama. Introduction to the modeling and analysis of complex systems. pages 90–94, 2015.