Community Segregation, Mean Field Theory, and the importance of Time on Friendship Formation

Karan Ruparell

This manuscript was compiled on June 7, 2023

This paper provides a comparison of mean field deterministic approximation with simulations on the formation of friendships in school, in the presecence of triadic closure. It also introduces a 'freezing rate' by which changes in the graph's edge set become less common over time. We also explore the effect of a time delay when schools form communities within themselves before being exposed to each other. We conclude that the mean field approximation is more appropriate for models that consider a freezing effect, as the randomness from edge change rates is diminished over time. For models that do not consider a freezing effect, the mean field approximation fails to show the existence of base fluctuations and the effect it has on community structure.

Social Networks | Mean Field Approximation | Scale Free Systems

1. Introduction

The aim of this paper is to determine the suitability of mean field deterministic approximation for models containing triadic closure. We will do so through a review of the current literature and an analysis of a new simulation feature and the mean field approximation. This paper will explore four key areas:

1. The Prevalence of scale-free distributions
2. Friendship structure within schools
3. Simulation Methods
4. A comparison of mean field theory approximations and other methods

This paper will then consider the impact of time on friendship formation. We introduce a freezing rate under which the probability of changes in the graph slows at a constant rate over time. We also introduce a delay by which schools are introduced to one another only after a set amount of time has passed, allowing them to form community structures before the introduction.

A. Definitions.

2: Wedge In complement to the definition of triadic closure, a triplet of nodes (a,b,c) is a wedge if the edge (a,b) and (b,c) are present but not (c,a).

1: Triadic Closure Triadic closure is the tendency for friends of friends to become friends, or equivalently the tendency for wedges to close. More formally, this means that $\mathbb{P}(A\, friends\, with\, B | A\, and\, B\, are\, friends\, with\, C) > \mathbb{P}(A\ friends\ with\ B)$. In the methods section we will discuss two ways of incorporating this tendency into simulations.

3: Homophily Homophily is a tendency for groups of individuals of the same 'type' to connect with each other, the opposite being heterophily. For s_i the average number of friendship agents of type i have with agents of the same type, and d_i the average number of friendships agents of type i have with people not of type i. We may define the *Homophily Index* as:

$$H_i = \frac{s_i}{s_i + d_i} \qquad [1]$$

When H_i increases with relative group size the friendship pattern satisfies *relative homophily*, and when H_i is larger than the relative group size w_i we say the group exhibits *inbreeding homophily*. We will consider the presence of inbreeding homophily in our experiments (1)

4: Mean field deterministic approximation Calculating the probability of each edge's existence and storing whether or not an edge exists in each time-step is computationally expensive. To have an easier way of monitoring the change in a variable, usually the total edges in a network (2), we use a mean field approximation. This works by replacing the system with a 1-body system, in this case the marginal probability of a random edge existing. We can then find the stable points of this system or calculate the sequence it takes from a starting point.

In the case of evolving networks in discrete time steps the mean field approximation is as follows: (2)

$$\langle A_{k+1}|X\rangle = \sum_{A_k \in S_n} F((A_k)P(A_k|X) \qquad [2]$$

Approximated as

$$\langle A_{k+1}|X\rangle \approx F(\sum_{A_k \in S_n} A_k P(A_k|X) = F(\langle A_k|X\rangle) \qquad [3]$$

Where F is the update distribution from one time step to the next.

Mean field theory is much simpler and easier to solve analytically, but it does not consider information about the graph such as modularity and degree distribution.

2. Literature Review

A. The Prevalence of scale-free distributions. Research on the topology of social networks has been common in the past decade, both in the case of schools and more widely (3–5).

One important debate is whether social networks commonly operate under a scale-free degree distribution. While empirically it has been found that social networks are often thought to be scale free, modelling studies have found that degree distribution of many networks can be modelled accurately using a Poisson, log-normal, and other distributions (6, 7). Even within scale free social networks, it is commonly agreed that cut-off points exist (8). Therefore nodes have a maximum degree they can take before being oversaturated. A point of concern regarding using a scale-free model is whether the cut-off point is sufficiently far enough away to be ignored. (5). This is also discussed by Artico in their 2020 paper (5). They point out that almost all empirically found networks are finite, and so networks derived from a scale-free degree distribution are indistinguishable from a cut-off scale-free network with a sufficiently distant cut-off point. Following this reasoning, they apply bootstrapping of the Solla Price preferential attachment algorithm to see that out of a set of over 4000 networks, approximately 64 percent could have come from a scale-free distribution. They chose to compare the networks to simulated data due to the degree dependency found within networks, where the degree of one node influences the degree of others. This is to say that they do not compare the data to a theoretical scale-free distribution, but models based upon scale-free distributions that incorporate degree dependence. Even so, they found that over a third of data does not fit a scale-free distribution, and so we must take further care and look more specifically at school friendships before claiming that the underlying distribution within schools is scale-free. For example, in best friend groups for social networks, R.I.M Dunbar shows that the cut-off points are too low to be ignored, suggesting that closer friendships may be an example where the scale-free distribution does not fit. (9) 1

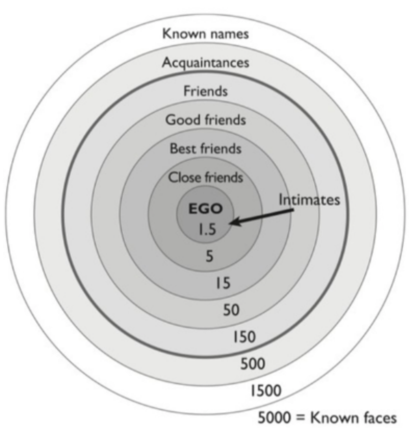

Fig. 1. Dunbar's Circle of Friends and expected friends in each

B. Friendship structure within schools. The considerable number of social networks available has allowed us to gain better insights into the relative frequency of different distributions, but it is also valuable to consider the structure of school friendships in particular.

Quintanilla et al.'s work showed that, for three elementary schools containing 100-400 children each, the degree distribution of friendships followed a Poisson distribution, with p values 0.107 and below. Erdos-Renyi graphs also produce Poisson degree distributions, and a plot of the final degree distribution under the method described in Grindrod et al.'s paper (2) closely resembles those from Quintanilla et al.'s, suggesting it may be a useful method for simulating school friendships.

Other papers also show that friendships in schools may not operate under the scale-free distribution. Empirical evidence by Gonzalez et al. (4) - summarised in figure 3a of their paper - shows that while the local clustering coefficient of friendships in secondary schools appears to distributed under a scale-free distribution, the degree distribution is not. In the case of Gonzalez's paper, the mode was 6-7 friends, which Dunbar1 and other sources also support(1, 9).

Under a more theoretic approach, Sun uses the Fokker-Plank equation to suggest that the log-normal distribution is in fact preferable when there is an optimum value, as there seems to be for friendship size. (10)

A key feature in social networks is the presence of communities. The demographic of people has been shown to affect the rate of triadic closure, and so community structure. (11) In schools this has been shown to include factors such as ethnic preferences (4), preferences based on which club-membership (12) and spatial preferences towards people in the same class (3). The spatial preferences suggest that people may be less likely to make friendship with people from different schools, all else being equal. However, it is unclear if this is only relevant when becoming friends with strangers or if the triadic closure rate is also affected by spatial preference.

Another mechanism we explore but has not yet been found in the existing literature is a decreasing rate of friendship changes. This is because as people get more familiar with the people in their schools they become increasingly less likely to make friends outside of their existing connections. This is both because they become used to their current setup and because they're more likely to have found their ideal group and number of friends. The changes this has on edge density and degree distribution are shown in figures 2 and 3.

C. Simulation Methods. There has been a substantial amount of literature on the growth of dynamic networks (13, 14), which describe how to maintain, predominantly scale free, distributions in growing social network models both in the cases of triadic closure (14) and without (13). There has also recently been more analysis on fixed-node systems (15), including in the case of fixed edges (16).

Grindrod and Higham produced early work on fixed-node community simulation (2, 17), which focused on Erdos-Renyi graphs, developing an algorithm with which to simulate changes in the network in relation to births δ, deaths $\tilde{\omega}$, and triadic closures ϵ. The papers use a Markov model for an efficient way of generating the probabilities of each edge existing. In their set up the only things that contribute to the probability of an edge's existence in the next time step is which edges exist in the current time-step.

They show a bistability, in which there are two stable points for the edge density of the graph, dependent on the ratio of the death rate with that of the triadic closure rate as $\delta \to 0$. Finally, they produce a mean field theory approximation in the special case of triadic closure and constant rates of birth, death, and closure where:

$$F(A_k) = (1 - \tilde{\omega})A_k + (1 - A_k)o(\delta\mathbf{1} + \epsilon A_k^2) \qquad [4]$$

They are then able to show that this has two stable points and one unstable point, where the two stable points are

$$\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{\tilde{\omega}}{\epsilon(n-2)}} + O(\delta)\ as\ \delta\ \to 0\ and\ \frac{\delta}{\gamma} \qquad [5]$$

In the case where there are two schools with no interaction, each has behaviour corresponding to the equations above with their own parameters.

They also attempted to adapt this method to preserve scale-free distributions, and found that this simulation method loses the information of the initial distribution. The simulations instead tend towards Poisson-like distributions (18). In their paper they use rules that depend upon specific cut-off points and the minimum or maximum degree of the two nodes involved in an edge but find that there is still a tendency towards Poisson distributions. This may be in part why recent papers(15, 16, 19, 20) have focused on a single potential edge change at each time step, unlike the batch updates initially proposed. These edge changes are designed with the degree distribution in mind, and only swap and introduce edges to preserve the degree distribution. This has become feasible because of the significant increase in computing power over the past decade. The more recent models are able to respond faster to changes in the network from growing edges, and allows the preservation of properties such as degree distribution (20).

One modern area of study in the role of triadic closure in social networks is in community segregation. Abebe et al. show in their recent paper (16) that triadic closure may have a positive effect on absolute network integration, where segregated communities merge, if and only if there is homophily, which is to say if and only if agents are more likely to form edges to agents within their community than others. Conversely however, triadic closure also has a negative relative effect on network integration in the case of homophily, which is to say that although the communities become more connected than they otherwise would have been, the communities also become more internally connected at a faster rate.

D. A comparison of mean field theory approximations and other methods. One limitation of mean field theory approximations is that they can often be difficult to solve analytically and require numerical approximation, as can be seen from Toyoizumi's 2006 paper "Generalisation of the Mean-Field Method for scale-free Distributions" (21) or Xing's later work generalising for exponential families (22). The sophistication of the proof required for this suggests that it may be too cumbersome to apply mean-field theory for distributions more complicated than those belonging to the alpha family, and less interpretable or amenable to changes. However, once cases of mean field approximation have been solved symbolically, as Toyoizumi and Xing have done, they can be used by any future people wishing to understand their specific system with different parameter values. This is not true for simulation methods, where although the simulation methods are easy to replicate the results are not necessarily generalisable to all other instances. In this way, mean field theory can provide us with stronger guarantees on a variety of general cases, but it is costly if not impossible to incorporate additional complications into the model that would otherwise be easy to incorporate through simulations. They also force us to reduce the problem to one with far less information, making it harder to deduce things like community structure or other high dimensional concepts.

Mean field theory also shows all solutions, including those that are not stable and thus not practically realisable. This is valuable as it can show us where to look for potential behaviours, which we may otherwise fail to find. Mean field theory also shows us the stability of steady states by checking starting points close to the solution. It however is not able to model the effect of randomness past this point, and the ability of trajectories to be knocked off onto different steady states by chance edge changes.

3. Method: Considering the impact of time on friendship formation

We will contribute to the research of Abebe et al. by considering the effect of triadic closure when there is a freezing effect added to the network, which makes the initial segregation more prominent.

The papers discussed suggest that there is unlikely to be scale-free degree distribution in school friendships, considering both empirical evidence from different year groups as well as mathematical and sociological evidence. As such we will not use models that preserve scale-free distributions (15).

One key factor that has not been discussed in the literature on discrete time step dynamic networks is whether the rates of birth, death, and triadic closure change over time. Here we provide simulations containing a constant freezing rate for all edge changes, crucially changing the stable points in (2) only up to $O(\delta)$, since the ratio $\frac{\gamma}{\epsilon}$ is preserved. This is relevant to the applicability of mean field theory, as it means that the previous stable points found are no longer likely to apply. We are, however, more likely to converge to the sequences predicted by mean field theory as the freezing rate also removes the background noise of edge changes.

We apply this change to the method described in (2). We believe this change is realistic as while students go through the school year they become busier and more entrenched in existing friendships, so are both less likely to make friends or to lose friends. Incorporating the limit to friendship sizes has been suggested by Higham in 2021 (19). One concern may be that the rate of triadic closure may not decrease at the same rate as meeting friends randomly. Alternatively, triadic closure may in fact increase as individuals interact more with existing friends. We believe however that because of the ideal number of friends triadic closure will still happen at a scale comparable to death rate. In recognition of this, we explore simulations that include an exponential decay of the birth and death rates. We use this to make suggestions on the importance of early action in avoiding community segregation (16). Currarini et al. have previously done similar work where they considered a mechanism by which each subsequent friend adds less value to the agent than previous friends, increasing the odds of starting friendships being maintained.(1) Our experiment adds value as it proposes a world where everyone, even those with less friends, become less willing or able to make friends over time.

Another change we make is adding a delay when introducing the schools to each other. We have a variable K corresponding to the number of time steps that occur before which the schools begin to interact with each other. This allows community structures to be built in the respective schools before introduction to the other school, and will allow us to evaluate the importance of early action when making groups interact with each other.

To isolate the effects of rate decay and delayed interaction, we treat the schools as being identical, and the internal friendships rates as identical to the external friendship rates. People have no preference between making and losing friends within their school and outside of it, conditioning for the number of mutual friends they have. While this seems an unlikely assumption, we have previously the importance of proximity in making friendships, this assumption allows us to better see the effects of our changes.

To make our model realistic we restrict to values of δ, ϵ, and γ that have a stable point corresponding to an average degree of 6-8 for each school. This corresponds to the ideal number of friendships.

Because of the resolution limit of modularity (23), it is not useful when looking at such small communities. As such we will not look at modularity, but instead use the inter-school density and average between-school density as a proxy for how disconnected the two schools are.

We choose our parameters to align closely with the results from Roy et al.'s recent work on the turnover in close friendships (24). It suggests that close friends, the group of ≈ 5 friends closest to you, have a turnover rate of 3-5 percent per year in people aged 17-24, over a three-year timespan. If we assume this is slightly lower than those in secondary school, as the data showed that turnover rate was lower for younger groups, our model describes a 2 year transition.

4. Results

Figure 2 2 shows that without any delay or freezing rate, we observe that the model correctly tends towards one of the solutions found under the mean-field equation, and eventually the edge density between schools becomes indistinguishable from those within schools. The mean field theory shows 10 real solutions, but we were only able to replicate 2, even when using at the other solutions as a starting point. The solutions that we could reproduce correspond to the stable points of an n1+n2 system with the chosen parameter rates. Under the parameters chosen, 600 steps are needed before convergence.

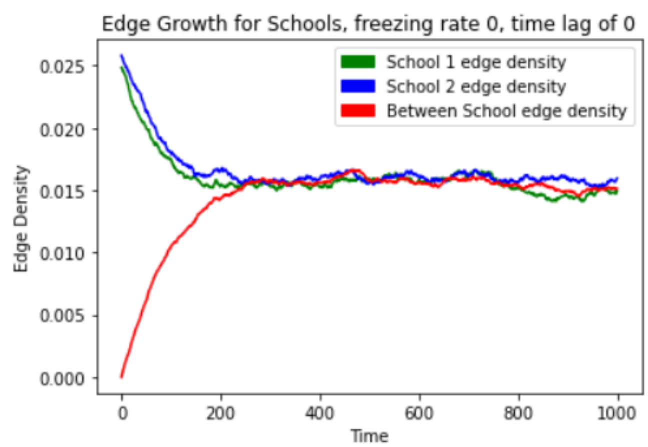

Fig. 2. Edge Densities from an Erdos-Renyi start with $\delta = 0.00015$ $\epsilon = 0.00045$ and $\gamma = 0.015$

When we add a freezing rate of only 1 percent as in Figure 3 3, such that after each time-step edges all parameters decrease by 1 percent, we find that the schools no longer become desegregated, but rather the connections between schools remain much sparser than those within. Within school densities stay close to the initial points, and the between-school densities settle considerably lower than the within school densities. Mean field theory suggests convergence to the same place as without the freezing rate, up to order δ, however in practice the freezing rate means that we do not converge there in any realistic time scale. As we can see in figure 2 even a 1 percent freezing rate convergence does not occur. Such a freezing rate also means that, under a scale-free starting point, the final degree distribution stays close to a scale-free one, not converging to a solution of the mean field equation as is predicted.

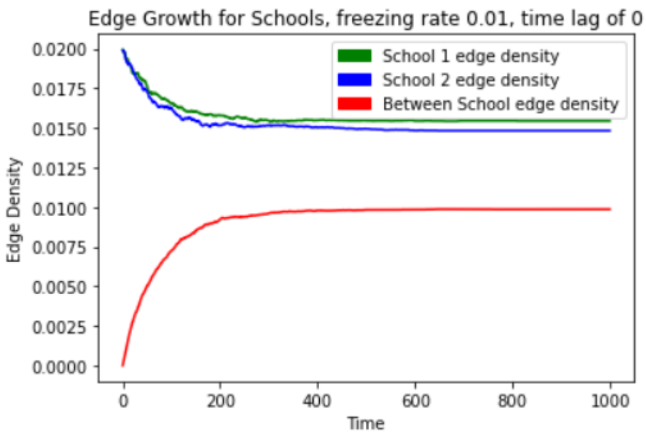

Fig. 3. A freezing rate of 1 percent resulting in entrenched community segregation.

The separation is still clear with a freezing rate of 0.5 percent, and only at a freezing rate of 0.1 percent do the simulations converge to the mean field approximation.

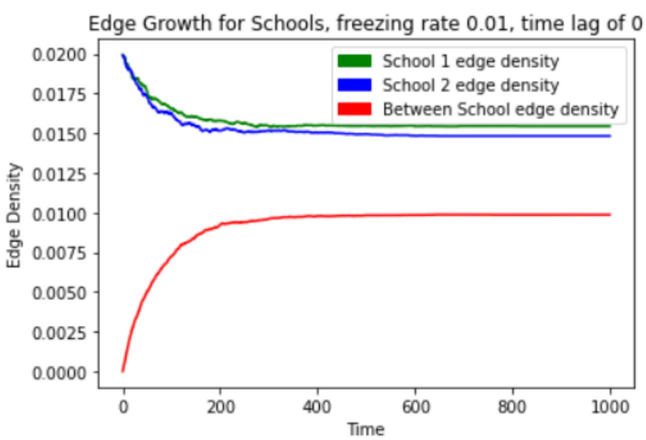

Fig. 4. A freezing rate of 0.1 percent resulting in no community segregation.

A delay in between-school interaction5 without a freezing rate has no effect on the final point of convergence. Before the interactions the two schools converge to one of their individual stable points - the stable points they would have as isolated schools - and then when interactions are introduced they converge to a stable point corresponding to a school of size $N_1 + N_2$.

Surprisingly, this does not change even when applying a delay for edge school interaction 6. Our initial hypothesis was that delaying interaction would mean that the friendships are effectively locked in before the two schools are able to interact, and that this would result in a lower between-school edge density even when the freezing rate was not high enough to stop convergence. This seems not to be true however, and while early action may counterfactually decrease the segregation when the freezing rate is strong enough to separate the group, on its own it does little to increase segregation itself.

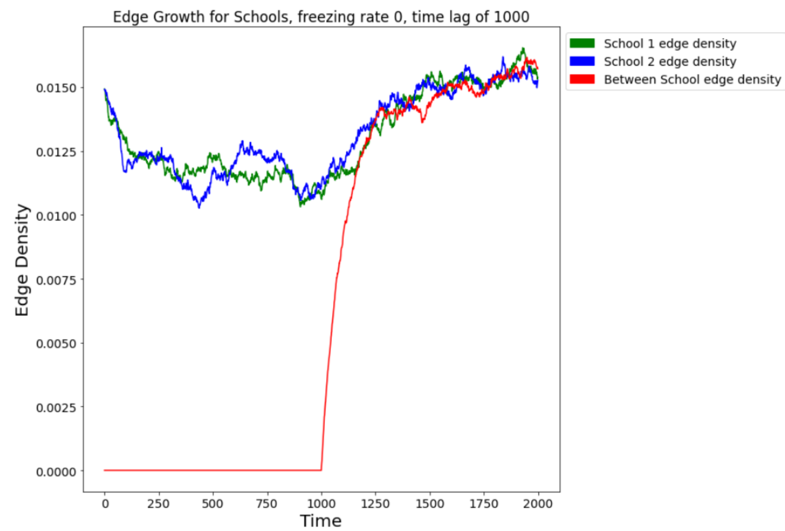

Fig. 5. A delay of 1000 time steps

Repeating this for smaller freezing rates and higher delays had no effect, and thus we conclude that delay rate has a much lower effect on relative segregation than freezing rate and initial conditions.

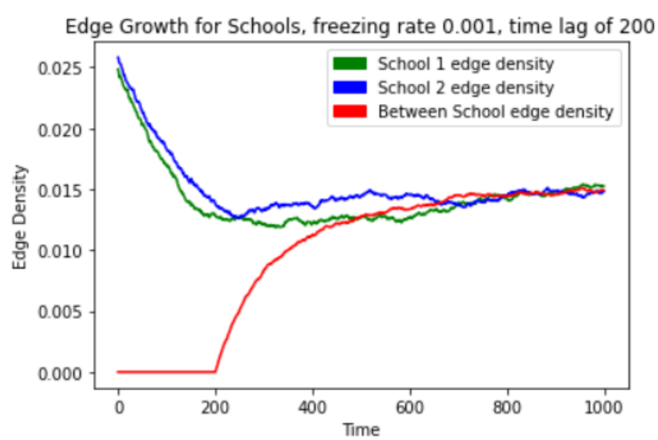

Fig. 6. A freezing rate of 0.1 percent and a delay of 200 time steps.

Separately, in following on from Grindrod et al.'s paper (18) we conducted experiments with multiple changes to the rates including: (i) birth and closure rates dependent on the geometric mean of the degrees affected, (ii) birth and closure rates dependent on the arithmetic mean of the degrees of the nodes in questions, (iii) birth and closure rates dependent on an Lp mean of the nodes in questions of the form $(\frac{d_i^p d_j^p}{\sum d_k^p}^{\frac{1}{p}})$. All of these resulted in a tendency towards a Poisson-like distribution.

5. Conclusion

In this work we have compared the potential distribution for friendships within schools and demonstrated the importance of reducing the freezing rate of friendships in promoting relative community integration. Our findings suggest that schools wishing to integrate their students with those from other schools would benefit from collaborative social events throughout the year, to reverse the freezing effect and increase the likelihood of more friendships forming. We have also shown evidence to suggest that early action may not be intrinsically important in decreasing relative community segregation, but only in so far as it increases overall network connectivity(2).

Further work is needed to establish the validity of a freezing rate, as this was outside of the scope of this study. A cut-off rate for the freezing rate may also be desirable, as the results from Roy et al. (24) suggest that there may be an additional underlying rate of turnover that exists regardless of how accustomed people are to their setting, with the middle aged cohort still losing friends at a rate of 4-6 percent over a 3 year span.

6. Acknowlegements

1. S Currarini, MO Jackson, P Pin, An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* **77**, 1003–1045 (2009).
2. MCP Peter Grindrod, Desmond J. Higham, Bistability through triadic closure. *Internet Math.* (2012).
3. R Huerta-Quintanilla, E Canto-Lugo, MD Viga-de Alva, Modeling social network topologies in elementary schools. *PloS one* **8**, e55371 (2013).
4. MC González, HJ Herrmann, J Kertész, T Vicsek, Community structure and ethnic preferences in school friendship networks. *Phys. A: Stat. mechanics its applications* **379**, 307–316 (2007).
5. I Artico, I Smolyarenko, V Vinciotti, E Wit, How rare are power-law networks really? *Proc. Royal Soc. A* **476**, 20190742 (2020).
6. H Fotuhi, A Amiri, MR Maleki, Phase i monitoring of social networks based on poisson regression profiles. *Qual. Reliab. Eng. Int.* **34**, 572–588 (2018).
7. P Van Mieghem, N Blenn, C Doerr, Lognormal distribution in the digg online social network. *The Eur. Phys. J. B* **83**, 251–261 (2011).
8. AL Barabási, Love is all you need: Clauset's fruitless search for scale-free networks. *Blog post available at https://www. barabasilab. com/post/love-is-all-you-need* p. 20 (2018).
9. R Dunbar, How many friends does one person need? in *How Many Friends Does One Person Need?* (Harvard University Press), (2022).
10. K Sun, Explanation of log-normal distributions and power-law distributions in biology and social science. *Dep. Physics, Univ. Ill. at Urbana-Champaign, Tech. Rep* (2004).
11. H Huang, J Tang, L Liu, J Luo, X Fu, Triadic closure pattern analysis and prediction in social networks. *IEEE Transactions on Knowl. Data Eng.* **27**, 3374–3389 (2015).
12. K Kafkas, ZN PERDAHÇI, M AYDIN, Ground truth in network communities and metadata-aware community detection: A case of school friendship network. *Alphanumeric J.* **9**, 49–62 (2021).
13. JSJHaKK Riitta Toivonen, Jukka-Pekka Onnela, A model for social networks. *Phys. A: Stat. Mech. its Appl.* (2006).
14. JI Ginestra Bianconi, Richard K. Darst, S Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks. *Phys. Rev. E* (2014).
15. SH Sara Ahmadian, The wedge picking model: A theoretical analysis of graph evolution caused by triadic closure and algorithmic implications. *J. Strateg. Innov. Sustain.* (2021).
16. JKBLAS Rediet Abebe, Nicole Immorlica, On the effect of triadic closure on network segregation. *Proc. 23rd ACM Conf. on Econ. Comput.* (2022).
17. DJH Alexander V. Mantzaris, Infering and calibrating triadic closure in a dynamic network. *Underst. Complex Syst.* (2013).
18. P Grindrod, DJ Higham, Models for evolving networks: with applications in telecommunication and online activities. *IMA J. Manag. Math.* **23**, 1–15 (2012).
19. KCZ Stefano Di Giovacchino, Desmond John Higham, A hierarchy of network models giving bistability under triadic closure. *Internet Math.* (2021).
20. A Asikainen, G Iñiguez, J Ureña-Carrión, K Kaski, M Kivelä, Cumulative effects of triadic closure and homophily in social networks. *Sci. Adv.* **6** (2020).
21. T Toyoizumi, K Aihara, Generalization of the mean-field method for power-law distributions. *Int. J. Bifurc. Chaos* **16**, 129–135 (2006).
22. EP Xing, MI Jordan, S Russell, A generalized mean field algorithm for variational inference in exponential families. *arXiv preprint arXiv:1212.2512* (2012).
23. S Fortunato, M Barthelemy, Resolution limit in community detection. *Proc. national academy sciences* **104**, 36–41 (2007).
24. C Roy, K Bhattacharya, RI Dunbar, K Kaski, Turnover in close friendships. *Sci. reports* **12**, 1–10 (2022).