Interacting Particle Systems on Dynamic and Scale-Free Networks

Fernley, John

Award date:
2021

Awarding institution:
University of Bath

Interacting Particle Systems on Dynamic and Scale-Free Networks

Fernley, John

Award date:
2021

Awarding institution:
University of Bath

Interacting Particle Systems on Dynamic and Scale-Free Networks

John Daniel Fernley

A thesis submitted for the degree of *Doctor of Philosophy*

University of Bath

Department of Mathematical Sciences

7th October 2020

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into [redacted]

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of work done in collaboration in [Fernley and Ortgiese, 2019], principally in Section 3.2 and Chapter 5. In this all calculations were performed by the author of the thesis (100%) while work in presentation of the content was equally share

Candidate's Signature [redacted]

Abstract

This thesis is concerned with the voter model and the contact process, two interacting particle systems in the sense of [Liggett, 1985]. For both systems, we are interested in the *termination* time of the process, called either consensus or extinction, and to bound this time in each case we shift our attention from the original process which is irreversible to a related reversible process. A reversible Markov chain is a simple random walk on a weighted graph with site-dependent stepping rates, and so there are many standard techniques for their analysis which we will apply.

The voter model is a classical interacting particle system modelling how global consensus is formed across a network, by local imitation. We analyse the time to consensus for the voter model when the underlying structure is a subcritical scale-free inhomogeneous random graph (in the sense of [Bollobás et al., 2007]). The reason that we focus on subcritical random graphs is that, as we will see below, the behaviour observed here cannot be captured by mean-field methods. Moreover, we generalise the model to include a 'temperature' parameter. The interplay between the temperature and the structure of the random graph leads to a very rich phase diagram, where in the different phases different parts of the underlying geometry dominate the time to consensus. We also consider a discursive voter model, where voters discuss their opinions with their neighbours. We find a different phase diagram for this discursive model in the subcritical case, and then begin to explore discursive voter model consensus and mixing on the supercritical network. Our proofs rely on the well-known duality to coalescing random walks and a detailed understanding of the structure of the random graphs.

Finally, we prove a phase transition for the contact process (a simple model for infection without immunity) on a homogeneous random graph that is initially Erdős-Rényi, but reacts dynamically to the infection to try to prevent an epidemic via *updating* in only the infected neighbourhoods, at constant rate. Under this graph dynamic, the presence of infection can help to prevent the spread and so many monotonicity-based techniques fail but analysis is made possible nonetheless via a forest construction.

Contents

Notation

We introduce a set of order notation for orders both *in probability* and *with polylogarithmic corrections*. For sequences of positive random variables $(X_N)_{N\geq 1}$ and $(Y_N)_{N\geq 1}$, we write $X_N = O_{\mathbb{P}}^{\log N}(Y_N)$ if

$$\exists K \in \mathbb{R} : \quad \mathbb{P}\left(X_N \leq Y_N (\log N)^K\right) \to 1$$

as $N \to \infty$. Similarly, we write $X_N = \Omega_{\mathbb{P}}^{\log N}(Y_N)$ if $Y_N = O_{\mathbb{P}}^{\log N}(X_N)$. If both bounds hold we write $X_N = \Theta_{\mathbb{P}}^{\log N}(Y_N)$. We write $X_N = o_{\mathbb{P}}^{\log N}(Y_N)$ if

$$\forall K \in \mathbb{R} : \quad \mathbb{P}\left(X_N \leq Y_N (\log N)^K\right) \to 1$$

as $N \to \infty$. Then, $X_N = \omega_{\mathbb{P}}^{\log N}(Y_N)$ if and only if $Y_N = o_{\mathbb{P}}^{\log N}(X_N)$.

We omit the $\log N$, writing $X_N = \Theta_{\mathbb{P}}(Y_N)$ etc., for the version of this definition with $K = 0$ and so disallowing polylogarithmic corrections.

Throughout we write $[N] = \{1, \dots, N\}$. For any graph G, we write $V(G)$ for its vertex set (which is typically $[N]$) and $E(G)$ for its edge set. The graph size $|G|$ is identified as the size of the vertex set $|V(G)|$.

Moreover, if $v, w \in V(G)$, we write $v \sim w$ if v and w are neighbours, i.e. if $\{v, w\} \in E(G)$. Then the degree of a vertex $v \in V(G)$ is $\mathrm{d}(v) := |\{w \in V(G) : v \sim w\}|$.

Abbreviations

ER Erdős-Renyi

CL Chung-Lu

NR Norros-Reittu

MNR Multigraph Norros-Reittu

SNR Simple Norros-Reittu

SCP Subtree Contact Process

SRW Simple Random Walk

CSRW Constant speed Simple Random Walk

VSRW Variable speed Simple Random Walk

SIR Susceptible-Infectious-Removed

SIS Susceptible-Infectious-Susceptible (or contact process)

CPEF Contact Process on an Evolving Forest

i.i.d. Independent and Identically Distributed

a.s. Almost Surely

w.h.p. With High Probability (or asymptotically almost surely)

w.l.o.g. Without loss of generality

List of Figures

Chapter 1

Introduction

The mathematical investigation of random graphs began in [Erdős and Rényi, 1959] which described the transition to connectedness for the uniform random graph with a fixed number of edges. Subsequently this partnership produced a great deal more, revealing this random object: in [Erdős and Rényi, 1960] the appearance of finite subgraphs is found quite generally to depend on a *threshold* order of edges; furthermore they find the size of the *giant component* when the asymptotic ratio of edges to vertices is at least $\frac{1}{2}$. Then, in [Erdős and Rényi, 1961], the first result was extended to see when this connectedness was robust to deletion of a finite number of vertices. Today we credit to Erdős and Rényi the related binomial model with independent edges which shows all these same features, although the precise definition of this model would be more properly attributed to [Gilbert, 1959].

Also in the 1950s, the graph theory language saw an explosion of application in an increasingly data-driven sociology and so "mathematicians began to think of graphs as the medium through which various modes of influence – information and disease in particular – could propagate" [Barabási et al., 2006].

We will investigate information and disease together within the framework of interacting particle systems [Liggett, 1985] which can simplistically model social interactions. The setting for these interactions is then a social network: nodes represent people and edges acquaintanceships. Of course, we cannot precisely prescribe every edge in a model of an extremely large and complex social network, and so we should think of these models by their random neighbourhoods in the "local weak limit". Global properties of the network can be discovered by fixing local properties in this way.

Locally and globally, power law degree distributions are "a common pattern seen in quite a few different networks" [Newman, 2010] and the network is in this case said to be

scale-free, a term coined in [Barabási and Albert, 1995] where they list an actor collaboration graph, the World Wide Web and the neural network of the worm *Caenorhabditis elegans* as examples. Social networks are prominently seen in this scale-free category and, therefore, we should take a network setting which is also scale free if we hope to prove properties which match phenomenologically (if not numerically) to the real systems with those degree distributions.

For information propagation, or more precisely opinion dynamics, the voter model is very popular and well-known. A significant factor in its popularity is that the voter model has a moment duality by time reversal to a system of coalescing random walkers, which we will exploit through much of this work.

In a sense the voter model has been considered since [Moran, 1958], as the Moran model can be seen as a voter model on a complete graph. It was introduced properly to a graph context when [Clifford and Sudbury, 1973] put it on a lattice, for the purposes of modelling territorial competition between species, but has since drawn considerable attention as a simple model for social dynamics. Shortly it was seen in the mathematics literature in [Holley and Liggett, 1975], in which they contrast discrete and continuous time versions. In the discrete time version on a finite graph we must rule out *periodicity* to guarantee consensus, which is less natural for modelling and so we restrict our attention to the continuous version.

The contact process, or susceptible-infectious-susceptible (SIS) infection, was introduced in [Harris, 1974] without application, for the sake of analysing its survival properties in $\mathbb{Z}^d$. Nonetheless, it is a very natural model for infection propagation in graphs which better fit that modelling problem. These social network models are often locally treelike (though the realism of this feature is disputed) and are always of course finite. Hence, we always see infection extinction but the transition between fast and slow extinction is frequently seen to match the transition between survival and extinction on the infinite random tree corresponding to the weak local limit [Bhamidi et al., 2019]. Further, we can access the fascinating double transition of [Pemantle, 1992], as seen in finite graphs by the work of [Mourrat and Valesin, 2016] on the d-regular network.

Initially, just as the Moran model can be thought of as an early version of the voter model which is either well mixed or on a large complete graph, infection models of this type were seen in their well mixed formulation in [Kermack and McKendrick, 1927]. They look at the susceptible-infectious-removed (SIR) version of the infection, in which new infections are made at rate proportional to the product of the number of infectious and susceptible vertices and *permanent* recoveries are made at rate proportional to the number of infectious vertices. The SIS infection which we analyse is the simplest

possible model, after SI in which vertices are permanently infected. There are many ways to add other stages in the family of compartmentalised models: for example SEIR inserts an "exposed" stage in which vertices are infected but not yet infectious. These are appropriate to modelling different infections, as discussed in [Bailey, 1975].

Applications of a certain type of mean-field technique, replacing the random network with a deterministic weighted complete graph, to analyse the contact process give the epidemic threshold of the inverse of the expected degree of a random neighbour [Pastor-Satorras and Vespignani, 2002, (4)]. That is, when the power law of the degree distribution is sufficiently large so as to make this expectation converge as the network size tends to infinity, for small enough infection rates we should find it impossible to see an epidemic. However this was found in [Chatterjee et al., 2009] to be incorrect, largely due to long survival around vertices of high degree. It is found in e.g. [Huang and Durrett, 2018, Lemma 2.4] that a star graph, with central degree k and k adjacent leaves, will on its own sustain a linear infection level for exponential time (in k), with high probability. Thus we are in a sense not seeing the fast mixing required for the efficacy of mean-field techniques and must be careful in analysis of the contact process. This is especially true in the presence of scale-free degree distributions, which produce on a network degrees polynomially large in the network size.

The voter model has no such issues with vertices of large degree in application of the same type of mean-field technique as carried out in [Sood et al., 2008], in the case when the edge density is sufficiently high to produce a giant component. However when the largest component is smaller order than the graph size, communication is very restricted and so this is when the technique loses validity. Hence, in the literature, the order of consensus time in this regime has so far not been established.

Chapter 2

Summary

This chapter contains all of the main results of this thesis, which will be proven through the other chapters. We also introduce and explain the definitions which are required to state the main results, before each of the two results sections.

Given a large set of people with labels $[N] := \{1, \dots, N\}$, and parameter $\gamma \in (0, 1)$, we can put a power law on these people

$$f : i \mapsto \left(\frac{N}{i}\right)^{\gamma}$$

and we see by counting that the weak limit $\frac{1}{N}\sum_{i=1}^{N} \delta_{f(i)} \overset{(\mathrm{d})}{\to} X$ has

$$\mathbb{P}\left(X > x\right) = \Theta\left(x^{-1-\frac{1}{\gamma}}\right)$$

as $x \to \infty$. Thus we deduce that $\tau := 1 + \frac{1}{\gamma} \in (2, \infty)$ is the tail exponent of the power law with these quantiles.

A random graph model is much more than simply a degree distribution, and there exists a wide variety of models for a scale-free network. Occam's razor has led more interest towards the simplest to define which is perhaps the Chung-Lu model [Chung and Lu, 2006]. This model on the set $[N]$ gives each node a *weight* setting that node's mean degree and then, for each pair of nodes $i \neq j$, generates the edge (i, j) independently with probability proportional to the product of the weights of i and j. This process discretises each real number weight in $f(i)$ to an integer degree, which converges weakly as $N \to \infty$ to a Poisson random variable.

If we further introduce a scaling constant $\beta > 0$ to control the global mean degree,

we obtain the inhomogenous random graph model with edge probabilities

$$p_{ij} := \frac{\beta N^{2\gamma-1}}{i^\gamma j^\gamma} \wedge 1 \tag{2.1}$$

which is exactly the *factor kernel* $(x, y) \mapsto \beta x^{-\gamma} y^{-\gamma}$ inhomogenous random graph seen in [Jacob et al., 2019]. We define here the relationship between a kernel and its inhomogenous random graph model.

Definition 2.0.1 (Inhomogenous Random Graph)**.** An Inhomogenous Random Graph (IRG), as in [Bollobás et al., 2007], will always here have ground space $(0, 1]$ with the Lebesgue measure, and vertex space $[N]/N \subset (0, 1]$. Then for any symmetric non-negative measurable kernel $\kappa : (0, 1]^2 \to [0, \infty)$, the IRG is the graph that independently features an edge between each distinct pair $x, y \in [N]/N$ with probability

$$\frac{\kappa(x, y)}{N} \wedge 1.$$

IRG models describe the *sparse* family of graph models, which are of interest for modelling social dynamics where the average person has a number of acquaintances which does not grow with the total number of people (given $\int_0^1 \kappa(\cdot, \mathrm{d}x) < \infty$). In the case $\gamma < \frac{1}{2}$, we have asymptotic equivalence between a great deal of IRG models.

Definition 2.0.2 (Asymptotic Equivalence)**.** Random graph sequences $(X_N)_N$ and $(Y_N)_N$ on $[N]$ are asymptotically equivalent if every valid sequence of graph sets $(E_N)_N$ has

$$\lim_{N\to\infty} (\mathbb{P}(X_N \in E_N) - \mathbb{P}(Y_N \in E_N)) = 0.$$

This is a very strong sense of equivalence, which as far as we seek to prove high probability behaviours (dependent on sequences of events with probability tending to 1) will allow arbitrary change between asymptotically equivalent definitions. In particular, the combination of [van der Hofstad, 2016, Theorem 6.19] and [van der Hofstad, 2016, Exercise 6.39] tell us that the Chung-Lu graph is equivalent to a certain parametrisation of the Norros-Reittu graph (which we define shortly) in this case.

Definition 2.0.3 (General Multigraph Norros-Reittu)**.** We parametrise Multigraph Norros-Reittu (MNR) by its vertex weights

$$\Lambda : [N] \to (0, \infty),$$

as in [Norros and Reittu, 2006]. The multigraph then has independently

$$\text{Pois}\left(\frac{\Lambda(i)\Lambda(j)}{\sum_{k=1}^{N}\Lambda(k)}\right)$$

edges between each pair $i, j \in [N]$, including pairs with $i = j$.

The Simplified Norros-Reittu (SNR) model is obtained from the multigraph model by *flattening*, i.e. removing loops (edges only incident to one vertex) and reducing all edge counts to 1 if they are greater. After this process, the one that agrees with the Chung-Lu parametrisation above is that with

$$\Lambda(i) = w(i) := \sum_{j=1}^{N} \beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \to \frac{\beta}{1-\gamma}\left(\frac{N}{i}\right)^{\gamma}$$

so we will use $\Lambda = w$ henceforth. Note, as well as the weight convergence we see above as $N \to \infty$, we see weak convergence of the degree $\mathrm{d}(i)$ to a Poisson distribution with mean of the limit weight. The limit mean degree over the whole graph, then, is $\frac{\beta}{(1-\gamma)^2}$. By flattening precisely this model, we arrive at the following model of a simple graph.

Definition 2.0.4 (Simplified Norros-Reittu)**.** The Simplified Norros-Reittu (SNR) graph, denoted G_N and with parameters $\beta > 0$, $\gamma \in [0, 1)$, is the simple graph with each edge $\{i, j\}$ independently present with probability

$$p_{ij} = 1 - \exp\left(-\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma}\right).$$

This is the graph model which we settle on, for being particularly easy to work with via its unflattened multigraph version, while still asymptotically equivalent to the Chung-Lu definition when $\gamma < \frac{1}{2}$. In fact, still when $\gamma < \frac{1}{2}$, this model is equivalent to a class of graphs which it is important to highlight.

Definition 2.0.5. Fix $\beta > 0$ and $\gamma \in \left[0, \frac{1}{2}\right)$. We say that a sequence of (simple) random graphs $(H_N)_{N\geq 1}$, where $V(H_N) = [N]$, is in the class $\mathcal{G}_{\beta,\gamma}$ if for any N there exists a symmetric array $(q_{ij})_{i,j\in[N]}$ of numbers in $(0, \frac{1}{2})$ such that each edge $\{i, j\}$, $i \neq j$, is present in H_N independently of all others with probability q_{ij}. Moreover, for $(p_{ij})_{i,j\in[N]}$ as in (2.1), we require that

$$\lim_{N\to\infty} \sum_{i\neq j} \frac{(p_{ij} - q_{ij})^2}{p_{ij}} = 0. \tag{2.2}$$

Every graph in the class $\mathcal{G}_{\beta,\gamma}$ in asymptotically equivalent to every other [van der Hofstad, 2016, Theorem 6.18], and it includes various well-known models of inhomogeneous random graphs additionally to the CL and SNR models defined above. These are "rank one" models as the matrix of CL probabilities (2.1) is a matrix of rank one.

In particular, it includes the Generalised Random Graph (GRG) with $q_{ij} = \frac{p_{ij}}{1+p_{ij}}$, which has the distribution of a particular configuration model conditioned to be simple [van der Hofstad, 2016, Theorem 7.18]. It follows therefore that this GRG model is precisely *uniformly distributed* on the set of random graphs with its degree distribution [van der Hofstad, 2016, Theorem 6.15] and so by the asymptotic equivalence we can see the whole class as uniform-type graphs. It is also possible, therefore, to justify these models by the "Principle of Indifference" [Keynes, 1921] in the absence of understanding of a complex network's detailed structure beyond its degree distribution.

Finally we comment that even in the case $\gamma > 1/2$ we can check the condition 2.2 on a subgraph using [van der Hofstad, 2016, Theorem 6.19] combined with the observation [van der Hofstad, 2016, Equation 6.8.13]. Thus we see that the SNR and CL networks do in fact have asymptotically equivalent induced subgraphs on $\left\{i \in [N] : i > N^{3/4}\right\}$ – in the bulk of the network, these models are still identical.

2.1 Voter Models

In what is now the standard model, voters can have one of two opinions and every vertex independently at rate 1 changes its opinion, forgetting its past opinion and copying that of one of its neighbours. More generally, we define the voter model as follows.

Definition 2.1.1 (Q-voter model)**.** Let O be the set of possible opinions and $Q = (Q(i,j))_{i,j\in[N]}$ be the generator of a continuous-time Markov chain on $[N]$. Given $\eta \in O^N$, define for $i \neq j \in V$,

$$\eta^{i\leftarrow j}(k) = \begin{cases} \eta(j) & \text{if } k = i \in V, \\ \eta(k) & \text{if } k \in V \setminus \{i\}. \end{cases}$$

The Q-voter model $(\eta_t)_{t\geq 0}$ is then the Markov process with state space O^N and generator

$$\mathcal{L}f(\eta) = \sum_{i=1}^{N}\sum_{j\neq i}^{N} Q(i,j)\left(f\left(\eta^{i\leftarrow j}\right) - f\left(\eta\right)\right).$$

In other words, at rate $Q(i,j)$ the voter i copies the opinion of voter j. If we

want to see this as a process on the graph $G = ([N], E)$, we will likely impose that $Q(i,j) > 0$ iff $\{i,j\} \in E$. Note that we could fill the off-diagonal elements of Q with any non-negative numbers and still obtain a Markov generator, so that there is no *a priori* reason to consider the Markov chain defined by Q. However it is natural to make further assumptions on the process defined by Q – we will always at least have that it is irreducible on the connected components of G so that this is a consensus-forming model.

We will also look only at voter models defined by a Q which is *reversible*. While this is necessary for tractability of the model, it is also natural for looking at local interactions on a locally treelike model: because an irreversible chain, by Kolmogorov's criterion, would need to have some nonzero flow of opinion around a cycle and cycles are typically large in the network, nonzero flow around such a cycle would require voters to see more than their finite neighbourhood.

One particular example of a valid Q is the standard model mentioned, where Q is the generator of the constant speed simple random walk on G. This has been studied in depth on $\mathbb{Z}^d$, see e.g. [Liggett, 1985], and typical questions study the structure and existence of invariant measures. When considered on a finite graph, the invariant measures become trivial and the main question is how long it takes to reach consensus. Inhomogenous random graphs are disconnected, so communication between components is impossible and we take the following definition.

Definition 2.1.2 (Consensus time)**.** Let $C_1, \ldots, C_k$ be the components of a graph G on the vertex set $[N]$. The *consensus time* is the first time that there is consensus on each component, i.e.

$$\tau_{\text{cons}} := \inf\{t \geq 0 \,:\, \eta_t|_{C_i} \text{ is constant for each } i \in [k]\} = \max_{i \in [k]} \tau_{\text{cons}}(C_i)$$

where $\tau_{\text{cons}}(C_i)$ will be occasionally used to denote consensus for the subgraph.

If the random walk with generator Q is irreducible on the connected components of G (which is automatic if it is reversible and has positive rates on every edge in G, and zero rates elsewhere), it can be seen that the consensus time for the voter model is also the first hitting time of any absorbing state. So, this is the only good definition for consensus when the process defined by Q is reducible.

Even the standard voter model has had no systematic treatment on a rank one scale-free network of the class $\mathscr{G}_{\beta,\gamma}$ from Definition 2.0.5. In the nonrigorous literature and so via a mean-field approach, [Sood et al., 2008] found consensus time of order N when $\tau > 3$ and of order $N^{\frac{2\tau-4}{\tau-1}}$ when $\tau \in (2,3)$. This model is very amenable to the mean-field

approach on a well-connected network so we largely believe these exponents are correct, but we will explore the cases of lower edge density in which their result does not hold.

Therefore we will analyse the standard voter model, also introducing a "temperature" parameter $\theta \in \mathbb{R}$ for which the standard model corresponds to $\theta = 0$.

Definition 2.1.3 (Classical voter model)**.** The *classical* voter model on $G = ([N], E)$ with temperature $\theta \in \mathbb{R}$ is the Q-voter model where for each distinct $i, j \in V$

$$Q(i,j) = \mathrm{d}(i)^{\theta-1}\mathbb{1}_{i\sim j}.$$

Compared to the standard voter model, this is an acceleration of the interaction based on the vertex degree: rather than resampling opinion at rate 1, the vertex v resamples its opinion at rate $\mathrm{d}(v)^\theta$. This extra parameter leads to interesting phase transitions in θ, where in the different phases different structural elements of the underlying random graphs dominate the consensus time.

The family of voter models described in Definition 2.1.3 can be called the "pull" family because agents pull opinions from their neighbours. There has been considerable research also into "push" families as in the original paper [Clifford and Sudbury, 1973] but we will not prove anything for these models. Instead, we look at the "push-pull" family also considered by [Moinet et al., 2018] and similar to the "oblivious" model of [Cooper et al., 2016].

Definition 2.1.4 (Discursive voter model)**.** The *discursive* voter model on $G = ([N], E)$ with temperature $\theta \in \mathbb{R}$ is the Q-voter model where for each distinct $i, j \in V$

$$Q(i,j) = \frac{1}{2}\big(\mathrm{d}(i)^{\theta-1} + \mathrm{d}(j)^{\theta-1}\big)\mathbb{1}_{i\sim j}.$$

In the pull models opinions were imitated and in the push models opinions were imposed. Hence in this model, the vertex v interacts at rate $\mathrm{d}(v)^\theta$, but then they "discuss" with a randomly chosen neighbour and agree on one opinion chosen at random between their two respective opinions. It is therefore a very natural model for social dynamics, more than simply a generator sum of the push and pull models.

2.1.1 Main results

All results in this section are stated on the Simplified Norros-Reittu network G_N of Definition 2.0.4, though they apply to other graph models in the class $\mathcal{G}_{\beta,\gamma}$ of Definition 2.0.5. Our first main theorem is on the expected consensus time whenever the parameters β, γ forbid the giant component.

Theorem 2.1.5. *Take $\beta > 0$, $\gamma \in \left[0, \frac{1}{2}\right)$ satisfying $\beta + 2\gamma < 1$ and initial conditions distributed as μ_u such that each initial opinion is an independent* Bernoulli(u) *random variable, for some $u \in (0,1)$. Then, for the classical voter model on G_N with parameter $\theta \in \mathbb{R}$, we have*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}}\left(N^c\right) \tag{2.3}$$

where the exponent $c = c(\gamma, \theta)$ is given as

$$c = \begin{cases} \gamma & \theta \geq 1, \\ \gamma\theta & \frac{1}{2-2\gamma} < \theta < 1, \\ \frac{\gamma}{2-2\gamma} & 0 \leq \theta \leq \frac{1}{2-2\gamma}, \\ \frac{\gamma(1-\theta)}{2-2\gamma} & \theta < 0. \end{cases}$$

The $\theta = 0$ classical dynamics on an averaged version of the graph are studied in [Sood et al., 2008] and they find $\Theta(N)$ expected time to hit *global* consensus whenever $\gamma < 1/2$. On the random graph, however, the consensus is componentwise (and note the largest component has $\Theta_{\mathbb{P}}\left(N^{\gamma}\right)$ vertices) and hence we find a faster polynomial order of consensus time $N^{\frac{\gamma}{2-2\gamma}} = N^{\frac{1}{2\tau-4}} = o(\sqrt{N})$. Note that we take an expectation just over the voter model dynamics, so that the expectation in (2.3) is still random (but depends only on the realization of the particular random graph).

We remark that the theorem shows that dominating contributions to the consensus time come from different parts of the random graph in the different regimes. On $\mathscr{C}(1)$, the component of vertex 1, which demonstrates the order $\Theta_{\mathbb{P}}\left(N^{\gamma}\right)$ of the largest component (and could be shown via [Janson, 2008, Remark 1.4] to in fact be with high probability the largest component) we have the following asymptotics.

Proposition 2.1.6. *In the same setting as Theorem 2.1.5, the subgraph consensus time on the largest component has*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}(\mathscr{C}(1))|G_N) = \Theta^{\log N}_{\mathbb{P}}\left(N^c\right), \quad \textit{where } c = \begin{cases} \gamma & \theta \geq 1, \\ \gamma\theta & \frac{\gamma}{1-\gamma} < \theta < 1, \\ \frac{\gamma^2}{1-\gamma} & 0 \leq \theta \leq \frac{\gamma}{1-\gamma}, \\ \frac{\gamma^2(1-\theta)}{1-\gamma} & \theta < 0. \end{cases}$$

Therefore, by comparison with Theorem 2.1.5, in the regime $\theta < 1/(2-2\gamma)$ we find $\mathscr{C}(1)$ is not the component that takes longest to reach consensus. Instead in the consensus time of Theorem 2.1.5, the dominating contribution comes from the consensus

time on a *double star* component, i.e. a tree component with two connected vertices of maximal lower degree, which exists with high probability.

We further comment that after restricting our attention to the largest component we still find an exponent diagram which is not monotonic, and so evidently the non-monotonicity is not created by different components with monotone individual diagrams as we might expect. Rather, the monotonicity is created by a competition between *mixing* and *space* for the coalescing dual model. For example, very negative θ values will slow the average speed of a walker (working against coalescence) but also tend the stationary distribution weakly towards the point measure on the highest degree vertex of a component (working towards coalescence).

In fact, we can slow the entire voter model by a factor to produce a constant average speed of leaving a vertex $\sum_{v\in\mathscr{C}(1)} q(v)\pi(v) = 1$, and then find $\mathscr{C}(1)$ exponents

$$\begin{cases} \gamma & \theta \geq 1, \\ \gamma\theta & \frac{\gamma}{1-\gamma} < \theta < 1, \\ \frac{\gamma^2}{1-\gamma} + \gamma\theta & 0 \leq \theta \leq \frac{\gamma}{1-\gamma}, \\ \frac{\gamma^2(1-\theta)}{1-\gamma} + \gamma\theta & \theta < 0. \end{cases}$$

which (after in this sense removing the difference in average speed) are monotone.

Next we consider the discursive model, where we have the following phase diagram.

Theorem 2.1.7. *Take $\beta > 0$, $\gamma \in \left[0, \frac{1}{2}\right)$ satisfying $\beta + 2\gamma < 1$ and initial conditions distributed as μ_u such that each initial opinion is an independent* Bernoulli(u) *random variable, for some $u \in (0,1)$. Then, for the discursive voter model on G_N with parameter $\theta \in \mathbb{R}$, we have*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}}(N^c)$$

where the exponent $c = c(\gamma,\theta)$ is given as

$$c = \begin{cases} \frac{\gamma}{2-2\gamma} & \theta \geq \frac{3-4\gamma}{2-2\gamma}, \\ \gamma(2-\theta) & 1 < \theta < \frac{3-4\gamma}{2-2\gamma}, \\ \gamma & 2\gamma \leq \theta \leq 1, \\ \frac{\gamma(2-\theta)}{2-2\gamma} & \theta < 2\gamma. \end{cases}$$

Unlike for the classical model, where large positive θ slowed down consensus when compared to the standard model $\theta = 0$, for the discursive model we see that large θ

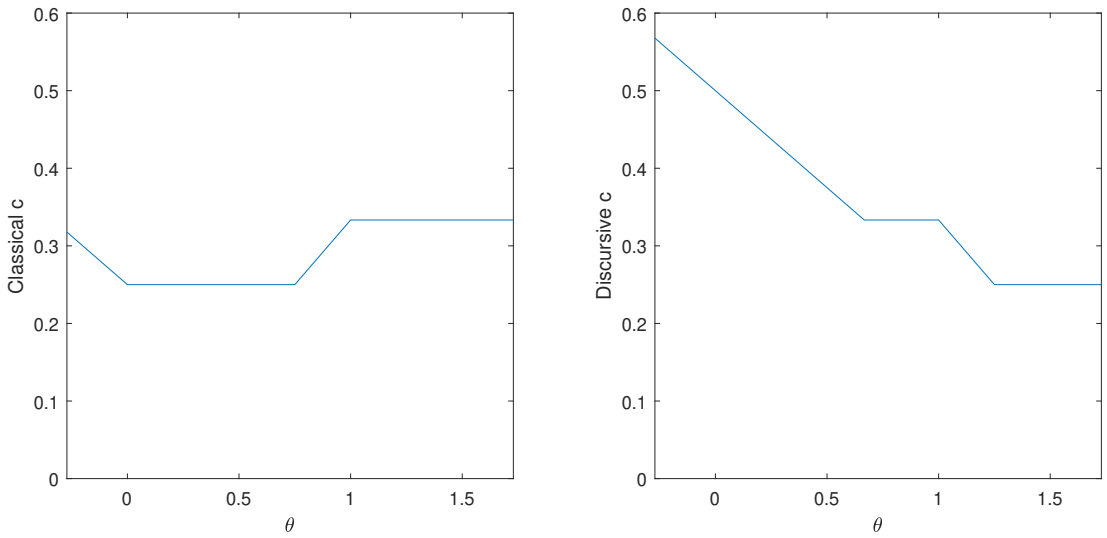


Figure 2.1: This figure shows the typical shapes of the exponents c in Theorems 2.1.5 and 2.1.7, by setting $\gamma = 1/3$. Somewhat surprisingly, for any subcritical (β, γ) parameters the function $\mathbb{c}(\gamma, \theta)$ is *not* monotonic in θ for the classical model. On the left we see that the standard model $\theta = 0$ is part of the fastest interval $\theta \in [0, 1/(2-2\gamma)]$. Conversely, essentially because the stationary distribution of the dual chain does not change with θ, the discursive model shows monotonicity in the exponent.

accelerates consensus by accelerating mixing: for each γ, $\mathbf{c}(\gamma, \theta)$ is non-increasing in θ. See also Figure 2.1 for an illustration. Again to understand in which regimes the large components dominate, we give the consensus order of $\mathscr{C}(1)$ with the discursive dynamic.

Proposition 2.1.8. *In the same setting as Theorem 2.1.7, the subgraph consensus time on the largest component has*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\mathrm{cons}}(\mathscr{C}(1))|G_N) = \Theta^{\log N}_{\mathbb{P}}\left(N^c\right), \quad \textit{where } c = \begin{cases} \frac{\gamma^2}{1-\gamma} & \theta \geq \frac{2-3\gamma}{1-\gamma}, \\ \gamma(2-\theta) & 1 < \theta < \frac{2-3\gamma}{1-\gamma}, \\ \gamma & 3 - \frac{1}{\gamma} \leq \theta \leq 1, \\ \frac{\gamma^2(2-\theta)}{1-\gamma} & \theta < 3 - \frac{1}{\gamma}. \end{cases}$$

The most obvious difference here is that $\mathscr{C}(1)$ makes a dominating contribution to the consensus order on G_N, as seen in Theorem 2.1.7, only for parameters θ in a an *intermediate* range $\theta \in [2\gamma, \frac{3-4\gamma}{2-2\gamma}]$ as opposed to in Proposition 2.1.6 where this was true for θ sufficiently large. We will see in the proofs that the consensus time in all regimes is dominated either by $\mathscr{C}(1)$ or a component of double star type, exhibiting slow mixing.

Remark 2.1.9 (Transitions in the power law)**.** *For illustration, we rephrase the main*

theorems by fixing θ and varying the tail exponent $\tau = 1+1/\gamma$. For the classical dynamics on G_N we obtain for $\theta \in \left(\frac{1}{2}, 1\right)$,

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}} \begin{cases} N^{\frac{1}{2\tau-4}} & \tau \leq 3 + 2\left(\frac{1-\theta}{2\theta-1}\right), \\ N^{\frac{\theta}{\tau-1}} & \textit{otherwise,} \end{cases}$$

and for the discursive dynamics with $\theta \in \left(1, \frac{3}{2}\right)$, this translates to

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}} \begin{cases} N^{\frac{1}{2\tau-4}} & \tau \leq 3 + 2\left(\frac{\theta-1}{3-2\theta}\right), \\ N^{\frac{2-\theta}{\tau-1}} & \textit{otherwise.} \end{cases}$$

In both these cases the consensus time on the largest component is dominant only for small τ. If $\theta \in (0, 1)$ then for the discursive dynamics we have that

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}} \begin{cases} N^{\frac{1}{\tau-1}} & \tau \leq 3 + 2\left(\frac{1-\theta}{\theta}\right), \\ N^{\frac{2-\theta}{2\tau-4}} & \textit{otherwise.} \end{cases}$$

In this case, one can see from the proofs that the asymptotics for the largest component dominate for large τ values.

For the discursive model in the context of a giant component, which exists if and only if $\beta + 2\gamma > 1$, we can prove a consensus time when $\theta \leq \frac{1}{\gamma}$. This at least covers all reasonable social modelling parameters: $\theta > 1$ would mean we expect agents with more contacts to also interact with a higher *proportion* of their contacts.

Theorem 2.1.10. *Take $\beta > 0$, $\gamma \in [0, 1)$ satisfying $\beta + 2\gamma > 1$ and initial conditions distributed as μ_u such that each initial opinion is an independent* Bernoulli(u) *random variable, for some $u \in (0, 1)$. Then, for the discursive voter model on G_N with parameter $\theta \leq \frac{1}{\gamma}$, we have*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}|G_N) = \Theta^{\log N}_{\mathbb{P}}(N).$$

With $\theta = 1$, this model is a timechange of the "link dynamics" of [Sood et al., 2008]. While they do not make a prediction for heterogeneous networks, because

$$\sum_{v \in [N]} q(v) = \sum_{v \in [N]} d(v) = \Theta_{\mathbb{P}}(N)$$

we have shown that consensus is expected in $\Theta^{\log N}_{\mathbb{P}}\left(N^2\right)$ link interactions.

When $\theta\gamma > 1$, then, we do not find polynomially tight bounds to establish the polynomial order; we discuss conjectures for the consensus time in this region in the introduction of Chapter 6. However, we do prove a statement for the time correlation of the voter model in this parameter regime. This follows from the following result for the mixing time which we state for its independent interest.

Definition 2.1.11 (VSRW)**.** The variable speed simple random walk on a graph is the Markov chain with, for every $i \neq j$, rate

$$Q(i,j) = \mathbb{1}_{i\sim j}.$$

We define the mixing time in Definition 4.1.2 by hitting of total variation distance $\frac{1}{4}$ from worst-case initial condition. We bound the mixing time for this walk on the SNR graph G_N, including the case $\gamma = 0$ though it would be straightforward to prove a faster order of mixing for $\gamma = 0$ using [Benjamini et al., 2014].

Theorem 2.1.12. *On the giant component of G_N with $\gamma \in [0,1)$ and β sufficiently large, the VSRW has mixing time*

$$t_{\text{mix}} = O_{\mathbb{P}}\left(\log^{17} N\right).$$

The VSRW has the same jump chain as the constant speed simple random walk (CSRW), which leaves every vertex at rate 1, but to our knowledge this chain also has no known mixing bounds on these graphs. Even in the $\gamma < \frac{1}{2}$ region of configuration model equivalence the existing results require further conditions on the minimum or maximum degree [Berestycki et al., 2018, Abdullah et al., 2012].

This theorem allows us to fairly easily deduce that the correlation decay in either voter model definition is seen over a polylogarithmic timeframe $\left[0, O^{\log N}(1)\right]$ for $\theta = 1$. In fact, there are monotonicities which allow this to be extended to $\theta \geq 1$.

Corollary 2.1.13. *Take $\gamma \in (0,1)$, β sufficiently large and initial conditions distributed as μ_u such that each initial opinion is an independent* Bernoulli(u) *random variable, for some $u \in (0,1)$. Define also the measure of time correlation*

$$C(t,v) = \frac{\text{Cov}\left(\eta_0(v), \eta_t(v)\right)}{u(1-u)}$$

recording correlation for the random process on a fixed graph.

Then for the discursive voter model on G_N with parameter $\theta \geq 1$, we have for any

$v \in [N]$ *that* $C(\cdot, v)$ *decreases monotonically from* $C(0, v) = 1$ *to*

$$C(T, v) = \begin{cases} \Theta_{\mathbb{P}}\left(\frac{1}{N}\right) & v \in \mathscr{C}(1) \\ \Omega_{\mathbb{P}}\left(\frac{1}{\log N}\right) & v \notin \mathscr{C}(1) \end{cases}$$

for any $T = \omega^{\log N}(1)$.

2.2 Contact Process

The contact process is an extremely simple Markovian model of an infection spreading on a graph: we record the state of the infection as $(\xi_t)_t$, where

$$\xi_t(\cdot) : [N] \to \{0, 1\},$$

and $\xi_t(v) = 1$ indicates that v is infected at time t, whereas $\xi_t(v) = 0$ indicates that v is healthy. For this to behave like an infection, we must account for both transmission and recovery and to obtain a Markov model each must happen at constant rate. By moving to a unitless relative measure of time, we can assume as is conventional that recoveries for each vertex occur independently at rate 1 and that transmissions of the infection occur for each edge independently at *relative* rate $\lambda > 0$.

Thus we can define the process by Markovian flipping at each vertex, where always

$$\{\xi_t(v) = 1\} \mapsto \{\xi_t(v) = 0\}$$

at rate 1, and the converse transition

$$\{\xi_t(v) = 0\} \mapsto \{\xi_t(v) = 1\}$$

is observed at rate $\lambda \sum_{w \sim v} \xi_t(w)$ – the number of infected neighbours of v at time t, multiplied by the relative infection rate. After every infection we return to susceptibility, and so the contact process will best model fast-mutating infections with low mortality.

The contact process is popular in part due to its self-duality [Liggett, 1999] which helps in many proof techniques. Here we write, for a set A, $\xi_t(A) := \sum_{a \in A} \xi_t(a)$.

Proposition 2.2.1. *On a static graph* $G = (V, E)$ *and* $A, B \subset V$ *the contact process has, for any fixed time* $t \geq 0$*,*

$$\mathbb{P}\left(\xi_t(A) > 0 \middle| \xi_0 \equiv \mathbb{1}_B\right) = \mathbb{P}\left(\xi_t(B) > 0 \middle| \xi_0 \equiv \mathbb{1}_A\right).$$

In particular, this allows easy transferral of results for the infection density of the process from full initial infection $\xi_0 \equiv 1$ to results about the survival probability of an infection spreading from a single vertex $v \in [N]$, with $\xi_0 \equiv \mathbb{1}_v$.

Just as for the voter model, this is a duality by time reversal and so it does in fact extend from static graphs to the non-adaptive dynamic we will consider, as this dynamic has the same distribution run backwards in time. Unfortunately, however, for the adaptive dynamic time reversal is not feasible and so we are forced to pick either full initial infection or single vertex. Hence we opt for the more practical case and consider the process from a generic initial infected vertex, say $\xi_0 = \mathbb{1}_1$, and we are interested in the size of the set of historically infected vertices or, more precisely, the occurrence of epidemic events where this set grows to linear scale.

Definition 2.2.2 (Epidemic events)**.** For the contact process $\xi : [0, \infty) \to \{0, 1\}^N$ define historical infection sets

$$I_t := \{v \in [N] : \exists s \in [0, t] : \xi_s(v) = 1\}$$

and the limit set $I_t \uparrow I_\infty$. Then, for any $\epsilon \in (0, 1)$ we can define the epidemic event

$$E_\epsilon^N := \{|I_\infty| > \epsilon N\} = \{\exists t > 0 : |I_t| > \epsilon N\}.$$

This is not all that we introduce to our contact process analysis. We analyse the contact process on an adaptive random graph: because the distribution of the graph (as a process in time) is made dependent on the infection we find the problem is complex enough on a homogenous graph and so drop the scale-free assumption.

If we shift focus to homogenous graphs by setting $\gamma = 0$, then the Simplified Norros-Reittu graph is asymptotically equivalent to the better-known Erdős-Rényi graph.

Definition 2.2.3 (Erdős-Rényi)**.** The Erdős-Rényi (ER) graph with parameter $\beta > 0$ is the simple graph with each edge $\{i, j\}$ independently present with probability $p_{ij} = \frac{\beta}{N}$.

Therefore, for clarity, we pick these edge probabilities β/N rather than the $\gamma = 0$ SNR probabilities $1 - e^{-\beta/N}$. It is hard to imagine that this will affect either the results or the proofs, though this will not be investigated.

For the sake of comparison, we will also look at the non-adaptive version of the dynamic. This can be described as in [Jacob et al., 2019] in the following construction.

Definition 2.2.4 (Non-adaptive dynamic Erdős-Rényi)**.** Take $N \in \mathbb{N}$. For each $x \in [N]$ let $\mathcal{U}^x = (U_n^x)_{n \geq 1}$ be an independent Poisson process of rate $\kappa > 0$ describing the update

times of x. Further for each distinct pair $x, y \in [N]$ let $\mathcal{C}^{x,y} = (C_n^{x,y})_{n\geq 0}$ be an i.i.d. sequence of Ber $\left(\frac{\beta}{N}\right)$ random variables. The dynamic graph $G_t = ([N], E_t)$ has edge $\{x, y\} \in E_t$ if and only if $C^{x,y}_{F^{x,y}(t)} = 1$, where

$$F^{x,y} : t \mapsto |\{s \in \mathcal{U}^x \cup \mathcal{U}^y : s \leq t\}| .$$

What this definition describes is *vertex updating*: independently and at constant rate, each vertex decides to *update* by completely moving its location on the network to an i.i.d. new location. This is done by simply deleting all incident edges and generating a number of new ones distributed as Bin $\left(N-1, \frac{\beta}{N}\right)$, with a uniformly chosen new set of neighbours. Note that because the vertex update events occur independently of the graph, the dynamic graph is Erdős-Rényi in distribution for all time.

Now we would like to build on this simple dynamic to create the adaptive dynamic. In this we would like to model people moving out of infected areas and so they will only move out of their current neighbourhood on the event that it contains the infection.

Definition 2.2.5 (Dynamic neighbourhood)**.** Given a dynamic graph history $(G_t)_{t\geq 0} = ([N], E_t))_{t\geq 0}$ we can define the dynamic neighbourhood

$$\Gamma_t(v) = \{w \in [N] : \{v, w\} \in E_t\}$$

as the set of neighbours of v at time t.

We can use this definition to introduce the adaptive graph dynamic, in which the edge $\{x, y\}$ is only refreshed by an update at incident vertex x only when x also happens to be adjacent to some infected vertex.

Definition 2.2.6 (Adaptive dynamic Erdős-Rényi)**.** The adaptive dynamic is constructed as in Definition 2.2.4, except that each $F^{x,y}$ is reduced to its adaptive version $F_A^{x,y}$ which is

$$F_A^{x,y} : t \mapsto |\{s \in \mathcal{U}^x : s \leq t, \xi_s(\Gamma_s(x)) > 0\}| + |\{s \in \mathcal{U}^y : s \leq t, \xi_s(\Gamma_s(y)) > 0\}| .$$

It is clear that $F_A^{x,y}$ is only a function of the graph history and hence this is a process we can define going forwards in time. It is not obvious from the above definition that the graph dynamic is Markovian – the clearer description for this property is that we thin from the previous updates $\mathcal{U}^x$ those times at which x had no infected neighbours.

With this adaptive definition, the dynamic graph becomes dependent with the infection and so despite the name "dynamic Erdős-Rényi", the adaptive graph $(G_t)_{t\geq 0}$ is *not*

Erdős-Rényi in distribution for any $t > 0$.

2.2.1 Main results

The first theorem describes a (λ, β) parameter region of small outbreaks for the infection with the nonstationary graph dynamic.

Definition 2.2.7 (Subcriticality)**.** Recalling the epidemic events of Definition 2.2.2 which looked at historically infected vertices, we say that a coupled infection process and dynamic graph is *subcritical* if for every $\epsilon > 0$

$$\lim_{N\to\infty} \mathbb{P}_1\left(E_\epsilon^N\right) = 0.$$

The subscript denotes that the initial infection set is $\{1\}$, so by exchangeability we are considering infection breaking out from an arbitrary single infected vertex.

We seek to prove a region of subcriticality using a stochastic upper bound, the subtree contact process of Definition 4.3.8, for which we can only forbid divergence when $\lambda\beta < e^{-1}$. This is a limit therefore on the size of the region that can be found.

Theorem 2.2.8. *For the contact process on an adaptive dynamic Erdős-Rényi graph we find if $\beta, \lambda, \kappa \geq 0$ satisfy*

$$\lambda\beta < 0.21 \tag{2.4}$$

then the sampled infection sets $(|I_\infty|)_{N\in\mathbb{N}^+}$ are tight in $\mathbb{N}^+$ and in particular the infection is subcritical.

We remark that if the dynamic is modified so that an update requires b infected neighbours and $b \geq 1$ is a fixed parameter, it is not hard to extend the above result to also apply in this case.

Definition 2.2.9 (Supercriticality)**.** Recalling the epidemic events of Definition 2.2.2 which looked at historically infected vertices, we say that a coupled infection process and dynamic graph is *supercritical* if there exists $\epsilon \in (0, 1)$ such that

$$\liminf_{N\to\infty} \mathbb{P}_1\left(E_\epsilon^N\right) > 0.$$

Also, we prove a theorem concerning survival of the infection to epidemic levels. We use the SIR model, in which recoveries are permanent, and by coupling to infection process and graph this bound is considerably easier to prove.

Theorem 2.2.10. *For the contact process on an adaptive dynamic Erdős-Rényi graph we find if $\beta, \lambda, \kappa \geq 0$ satisfy*

$$\frac{\lambda\beta}{1 + 2\kappa + \lambda} > 1,$$

then the infection is supercritical.

Not as a showcase result, but for comparison, we also look at the non-adaptive dynamic where every vertex updates at rate κ regardless of the infection level of its neighbourhood. With the stationary dynamic, introducing the SCP is not necessary and so we can extend all the way up to the mean-field limit $\{\lambda\beta < 1\}$.

Theorem 2.2.11. *For the contact process on a non-adaptive dynamic Erdős-Rényi graph we find if $\beta, \lambda, \kappa \geq 0$ satisfy:*

(a) $\lambda\beta < 1$ and κ sufficiently large, then the sampled infection sets $(|I_\infty|)_{N \in \mathbb{N}^+}$ are tight in $\mathbb{N}^+$ and in particular the infection process is subcritical;

(b) $\lambda\beta > 1$ and κ sufficiently large, then the infection process is supercritical.

Here the more tractable mean-field model can be identified as $\kappa = \infty$ and so the appearance of a large κ condition was predictable. We can simulate an answer for the epidemic regions of each model using the contact process on an evolving forest (CPEF) that we will later define in Definition 4.3.5 (with natural modifications in the non-adaptive case), and we see then that perhaps this large κ condition is in fact only technical for the region of supercriticality, but still the critical line does curve towards a value less than 1 for the $\kappa = 0$ model. This is the critical value of [Nam et al., 2019, Corollary 3], a result which tells us that the $\kappa = 0$ critical value is asymptotic to $\frac{1}{\beta}$ as $\beta \to \infty$ and so suggests this critical line will straighten in that limit.

Each simulation of Figure 2.3 generates a version of the local tree environment seen by the infection; when averaging over the random graph in this way it seems that the non-adaptive influence of κ is mean-field for κ sufficiently large. There is undoubtedly some curve in the transition to the static case $\kappa = 0$ which is more mysterious.

Predictably, for the adaptive dynamic, we see in Figure 2.2 that the region where we can show survival by a contained SIR infection is not the whole survival region. The complete condition appears similar to that of Theorem 2.2.10 in that it has a linear boundary and requires $\lambda\beta > 1$, but the slope of the boundary is of course quite different.

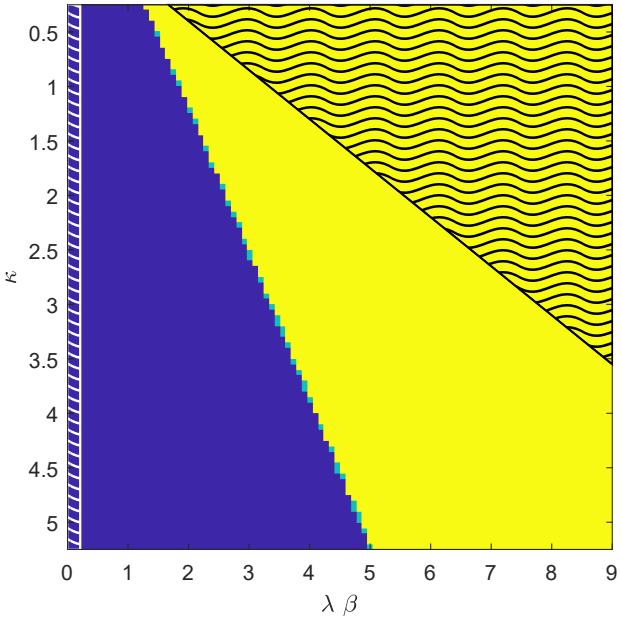


Figure 2.2: For the adaptive dynamic with mean degree $\beta = 10$, after 10^5 samples in the yellow parameter regions we have two standard deviations of confidence for supercriticality, which we approximately identify as 98% confidence. In the blue regions we have the same for subcriticality. Above the superimposed black line is the epidemic region proved in Theorem 2.2.10, whereas to the left of the white line is the region of Theorem 2.2.8

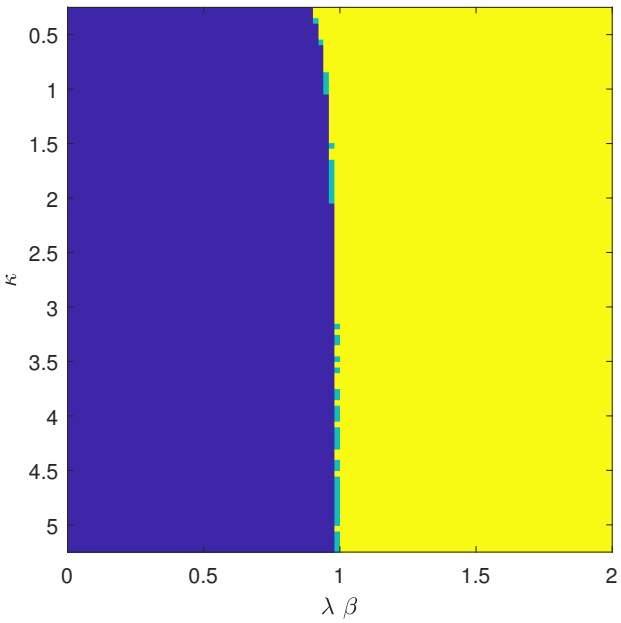


Figure 2.3: For the non-adaptive dynamic with mean degree $\beta = 10$, after 10^6 samples in the yellow regions we have two standard deviations (roughly 98%) of confidence for supercriticality and in the blue for subcriticality. Here we see in contrast to Figure 2.2 that the κ plays almost no role.

Chapter 3

Networks

To prove results for the behaviour of interacting particle systems on the SNR network we must understand its structure, which in many instances means that we must prove a statement about the network which has no explicit reference to the process that we are ultimately analysing. This chapter goes through all of the necessary structural properties, which often cannot be found in the literature even if similar properties are proven for similar graphs.

For example, while [Bollobás et al., 2007] gives general upper bounds for the diameter of IRGs with finitely many vertex types or a bounded kernel, the SNR graph fits neither of these conditions. Hence in section 3.3 we use their result to extend a subgraph IRG with finitely many types and obtain the diameter bounds we require for our main theorems, which include the construction of paths through vertices of low degree.

Section 3.1 is of particular importance for the construction of components and component subgraphs through many of the proofs. The elegance of this local construction is the large part of the appeal of working with Norros-Reittu graphs, and while our version of the construction is similar to that of [Norros and Reittu, 2006] we also retain the thinned edges which would create a cycle in the network. The original construction builds a spanning tree while ours builds the full graph, and thus we are able to continue the local construction past the initial window where the local graph is a tree up to the generation of complete components.

We first observe that the network models we have chosen show a stark phase transition in the emergence of a giant component, which is very relevant for the behaviour of interacting systems on the networks.

Definition 3.0.1 (Giant component)**.** Let $(H_N)_N$ be a sequence of random graphs on

$[N]$ and for each

$$\mathscr{C}_{\max}(H_N)$$

denotes the largest connected component subgraph (or the joint largest having smallest minimum element in the event of a tie). We say that the network model $(H_N)_N$ has a giant component if

$$|\mathscr{C}_{\max}(H_N)| = \Theta_{\mathbb{P}}(N)$$

and otherwise that it does not have a giant component.

Using the IRG framework of Definition 2.0.1, there is a technical condition to check.

Definition 3.0.2 (Graphical kernel)**.** The sequence of kernels $(\kappa_N)_N$ on $(0,1]^2$ with the Lebesgue measure ν is graphical iff

1. ν-a.e. for $(x,y) \in (0,1]^2$, $(x_N, y_N) \to (x,y) \implies \kappa_n(x_N, y_N) \to \kappa(x,y)$.
2. κ is continuous ν-almost everywhere on $(0,1]^2$.
3. $\kappa \in L^1((0,1] \times (0,1], \nu \times \nu)$
4. If H_N denotes the random graph model defined by κ_N in Definition 2.0.1, this model has expected edge count (half the total degree)

$$\frac{1}{N}\mathbb{E}(|E(H_N)|) \to \frac{1}{2}\int_0^1 \int_0^1 \kappa(x,y)\mathrm{d}x\mathrm{d}y.$$

We will identify these theorems heuristically as corresponding to the situations when the offspring mean for the local weak Galton-Watson limit has offspring mean larger or greater than 1, but before this explanation we can state the clear form.

Theorem 3.0.3 ([Bollobás et al., 2007] Theorem 3.1(i))**.** *If we define the operator*

$$T_\kappa f(x) = \int_0^1 \kappa(x,y) f(y)\mathrm{d}y$$

for some graphical kernel κ, with norm

$$\|T_\kappa\| = \sup\{\|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1\}$$

and H_N denotes the IRG with kernel κ, then we find

1. $\|T_\kappa\| \leq 1 \implies |\mathscr{C}_{\max}(H_N)| = o_{\mathbb{P}}(N)$,

2. $\|T_\kappa\| > 1 \implies |\mathscr{C}_{\max}(H_N)| = \Theta_{\mathbb{P}}(N)$.

Thus this theorem has precisely separated the regimes where the giant component exists and where it doesn't. In fact further, this theorem partitions the parameter values entirely into those where the linear component is w.h.p. present and those where all the components are sublinear w.h.p..

Corollary 3.0.4. *When $\gamma < 1/2$, the SNR model G_N of Definition 2.0.4 has a giant component iff $\beta + 2\gamma > 1$.*

Proof. Existence of a giant component is only a claim *with high probability*, and so we can in this case use another model in the class 2.0.5 that we have already discussed: the Chung-Lu model. This is the IRG with kernel

$$\kappa(x, y) = \beta x^{-\gamma} y^{-\gamma} =: \psi(x)\psi(y)$$

so as in [Bollobás et al., 2007, Section 16.4] because $T_\kappa f = \left(\int_0^1 f(x)\psi(x)\mathrm{d}x\right)\psi$ we have

$$\|T_\kappa\| = \|\psi\|_2^2 = \int_0^1 \beta x^{-2\gamma}\mathrm{d}x = \frac{\beta}{1-2\gamma}$$

which exceeds 1 if and only if $\beta + 2\gamma > 1$. □

Corollary 3.0.5. *When $\gamma \geq 1/2$ and $\beta > 0$, the SNR model G_N of Definition 2.0.4 has a giant component.*

Proof. The function $x \mapsto e^{-x} - x$ changes sign over $(0, 1)$, so by the Intermediate Value Theorem we can take $\epsilon \in (0, 1)$ such that $\epsilon = e^{-\epsilon}$.

We also have the bound for any $x > 0$

$$1 - e^{-x} > x(1 - x)$$

so that for $i, j \in [N]$ such that $\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} < \epsilon$ we can infer for the SNR edge probabilities

$$p_{ij} = 1 - \exp\left(-\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma}\right) > (1 - \epsilon)\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma}.$$

Otherwise, if $\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \geq \epsilon$, we have instead

$$p_{ij} = 1 - \exp\left(-\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma}\right) \geq 1 - \exp(-\epsilon) \geq (1 - \epsilon)\left(1 \wedge \beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma}\right)$$

where the last inequality follows from recalling $\epsilon = e^{-\epsilon}$. Combining both bounds, we have shown that the SNR model dominates the Chung-Lu model after edge percolation of the latter with retention probability $1 - \epsilon$. Because $\gamma \geq \frac{1}{2}$, the Chung-Lu model has operator norm

$$\|T_\kappa\| = \int_0^1 \beta x^{-2\gamma} \mathrm{d}x = \infty$$

and by [Bollobás et al., 2007, Corollary 3.3] this model has a giant component for any positive edge retention probability.

Then we have that we can build the SNR model by adding edges to a percolated model which already has a giant component, and so we conclude that the SNR model too has a giant component. □

3.1 Coupling with a Branching Process

We adopt a useful concept as stated in [van der Hofstad, 2020]. Write $\mathscr{G}_*$ for the space of graphs *modulo graph isomorphisms* rooted at some vertex $\circledcirc$, i.e. the space of equivalence classes of rooted graphs. Then put a metric on this space: for two rooted graph isomorphism classes $G_1, G_2 \in \mathscr{G}_*$ we can define each's (isomorphism class of a) ball around the root of radius r as $B^{(G_1)}(\circledcirc, r)$ and $B^{(G_2)}(\circledcirc, r)$ respectively, and then

$$R := \sup_r \left\{ B^{(G_1)}(\circledcirc, r) = B^{(G_2)}(\circledcirc, r) \right\}$$

induces the distance

$$d_{\mathscr{G}_*}(G_1, G_2) = \frac{1}{1+R}.$$

Definition 3.1.1 (Local weak convergence in probability)**.** Let H_N be a random graph on $[N]$ and $\circledcirc_N \sim U[N]$ an independently uniform root. We say $(H_N, \circledcirc_N)$ converges locally weakly in probability to the random rooted graph $(H, \circledcirc) \in \mathscr{G}_*$ if for every bounded and continuous function $h : \mathscr{G}_* \to \mathbb{R}$ we have

$$\mathbb{E}\left(h(H_N, \circledcirc_N) \middle| H_N\right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(h(H, \circledcirc)\right).$$

We will not require this concept in the proofs until Chapter 7 but we state it here because this section is a construction of the local weak picture $(G, \circledcirc)$ for our network model G_N.

The reason for preferring the SNR model is the close relation with the standard multi-graph Norros-Reittu (MNR) model. Recall that the SNR model with edge probability

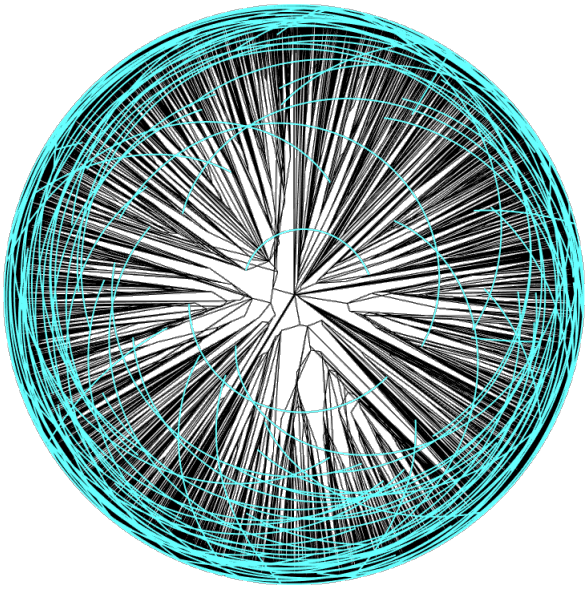


Figure 3.1: If we take $\gamma < \frac{1}{2}$ and $\beta > 1 - 2\gamma$, on the event that our root vertex is in the giant component, the underlying thinned Galton-Watson trees grows exponentially from the root, shown here as black edges. As it grows, after discovering $\Omega_{\mathbb{P}}(\sqrt{N})$ vertices, we encounter cycle edges, shown here in cyan, which give the giant component its nonzero surplus.

as in Definition 2.0.4 is then obtained from MNR by constructing it with weights

$$w(i) := \sum_{j=1}^{N} \beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \sim \frac{\beta}{1-\gamma} \left(\frac{N}{i}\right)^{\gamma}, \tag{3.1}$$

and then collapsing all multi-edges to simple edges and deleting the loops.

The MNR model is particularly nice because it allows for an exact coupling with a two-stage Galton-Watson process, with thinning and cycle creation. Our construction here extends the coupling introduced in [Norros and Reittu, 2006] (see also [van der Hofstad, 2016]) by also keeping track of the number of edges, so that we can also control when we create cycles.

Define the *mark distribution* to be the random variable M on $[N]$ which chooses a vertex biased proportional to its weight

$$\mathbb{P}(M = m) \propto w(m) \mathbb{1}_{m \in [N]} \propto m^{-\gamma} \mathbb{1}_{m \in [N]}$$

so that if W_N is the empirical weight distribution in the network, the weight of a typical

neighbour in our local picture will be simply the size-biased version of W_N, denoted W_N^*

$$w(M) \stackrel{(\mathrm{d})}{=} W_N^*.$$

Fix $k \in [N]$, we now describe the (marked) branching process that describes the cluster exploration when started from a vertex k. To describe the branching process, we label the tree vertices using the standard Ulam-Harris notation, in particular we denote by $\emptyset$ the root of the tree, by 1 the first offspring of the root, by 11 the first offspring of tree vertex 1 etc. We will write $v < w$ if v comes first in the breadth-first ordering of the tree, i.e. vertices are first sorted according to length and then according to lexicographical ordering if the lengths are the same.

For the root of the branching process, we define

$$M_\emptyset = k, \; X_\emptyset \sim \mathrm{Pois}\left(w(k)\right).$$

Next, we define independent random variables $(X_v)_{v\neq\emptyset}$ in two stages: we first choose marks $(M_v)_{v\neq\emptyset}$ which are i.i.d. with the same distribution as M. Then, conditionally on M_v, let $X_v \sim \mathrm{Pois}\left(w(M_v)\right)$. where we write $\mathrm{Pois}(Y)$ for the mixed Poisson law with random mixing parameter Y.

Moreover, if we take X_v to be the number of children of vertex v (if it exists in the tree), this construction can be used to define a (marked) Galton-Watson tree $\mathcal{T}^k$ (where only the root has a different offspring distribution).

To obtain the cluster at k in G_N from $\mathcal{T}^k$, we introduce a thinning procedure. We set $\emptyset$ to be unthinned and then explore the tree in the breadth-first order described above and thin a tree vertex w if either one of the tree vertices in the unique path between $\emptyset$ and w has been thinned or if there exists an unthinned $v < w$ with $M_v = M_w$.

Now, denote for $i \in [N]$, $X_v(i)$ to be the number of children of v with mark i. If v and w are unthinned tree vertices, then we define

$$E(M_v, M_w) = \begin{cases} X_v(M_w) & \text{if } v < w, \\ X_w(M_v) & \text{if } w \leq v, \end{cases} \tag{3.2}$$

so that edges between M_v and M_w are only generated at the exploration of the earlier of the two, in the breadth first ordering.

We can define the multigraph $\mathcal{T}^k_{\mathrm{thin}}$ by specifying that the vertex set is $\{M_v : v \text{ unthinned}\}$ and the number of edges are given by $(E(M_v, M_w) : v, w \text{ unthinned})$.

Similarly, we can define a forest $(\mathcal{T}^1, \mathcal{T}^2, \ldots, \mathcal{T}^n)$ of independent trees constructed

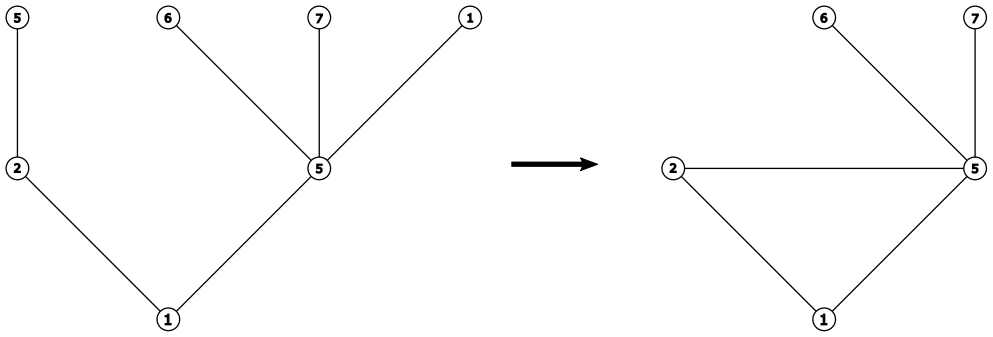


Figure 3.2: On the left a realisation of the labelled Galton-Watson tree $\mathcal{T}^1$ and on the right the resulting random graph. Note that the number of edges between 2 and 5 are determined by the number of children of type 5 of the first child of the root.

as above, where the root of the kth tree has mark k. Then, we can define the same thinning operation as above, starting in the tree $\mathcal{T}^1$ and going to the next when the algorithm terminates, where now also the roots of the trees may be thinned if their label has appeared in a previous tree. If we define the edges as in (3.2), then we obtain a multigraph $(\mathcal{T}^1, \mathcal{T}^2, \dots, \mathcal{T}^n)_{\text{thin}}$ with vertex set $\{M_v : v \text{ unthinned}\} = [N]$ and the number of edges between i and j given as $E(M_v, M_w)$, where v and w are the unique unthinned vertices v, w with $M_v = i$ and $M_w = j$.

With this construction, we have the following proposition.

Proposition 3.1.2. *Let G_N^{MNR} be a realization of a Norros-Reittu mulitgraph. For any fixed vertex $k \in [N]$, we have for the component $\mathscr{C}(k)$ in G_N^{MNR} containing k,*

$$\mathscr{C}(k) \stackrel{d}{=} \mathcal{T}^k_{\text{thin}}.$$

Moreover,

$$G_N^{\text{MNR}} \stackrel{d}{=} (\mathcal{T}^1, \mathcal{T}^2, \dots, \mathcal{T}^n)_{\text{thin}}.$$

This proposition can be proved in the same way as Prop. 3.1 in [Norros and Reittu, 2006]. The only difference is that we explicitly keep track of the number of edges.

Remark 3.1.3. Thinning and creation of cycles. *Note that by construction, we only create edges that lead to cycles if there are tree vertices v, v', w such that v' is a child of v, but in the breadth-first order $v < w < v'$, v and w are unthinned and such that $M_w = M_{v'}$, see Figure 3.2 for an example. The reason for this is that the number of edges between M_w and M_v is determined by looking at the types of children of the first vertex in breadth-first order. As a consequence, by this procedure we do not create edges*

between different components of $\mathcal{T}^i$. Moreover, if an unthinned tree vertex v has children $v'_1, \dots, v'_\ell$ with $M_v = M_{v'_i}$, then this leads to ℓ self-loops.

Remark 3.1.4. *Note that for the second construction, if the root of the kth tree $\mathcal{T}^k$ is not thinned, then any vertex in the tree with root k that receives mark $j \leq k$ will be thinned. So to get a stochastic upper bound on the number of vertices and their degrees in the component $\mathscr{C}(k)$, we can replace $\mathcal{T}^k$ by $\mathcal{T}^k_k$, where the marks are chosen independently with distribution*

$$M_k \stackrel{(\mathrm{d})}{=} \begin{cases} M & \text{if } M > k, \\ \dagger & \text{otherwise.} \end{cases}$$

Then, the offspring distribution is $\mathrm{Pois}(W^*_{N,k})$ *with* $W^*_{N,k} \stackrel{(\mathrm{d})}{=} w(M_k)$*, where we set* $w(\dagger) = 0$*. The error in this upper bound comes from thinning within $\mathcal{T}^k$ and also that thinned vertices are included as leaves of zero weight rather than simply being removed.*

This local weak picture of the branching process combined with results in [Kesten and Stigum, 1966, Bingham and Doney, 1974] shows that the finite radius neighbourhoods $|B(o,r)|$ grow exponentially in r when $\gamma < \frac{1}{2}$. In the case $\gamma > \frac{1}{2}$, we can apply [Davies, 1978] to see that then $\log |B(o,r)|$ grows exponentially in r, i.e. growth is *doubly exponential*.

3.2 Structural Results for Subcritical Random Graphs

In this section, we collect some of the structural results on subcritical inhomogeneous random graphs that we will need later on in Chapter 5. Some of these results are known, but as the literature on subcritical inhomogeneous random graphs is less developed than for supercritical random graphs, we have to prove the more specialised ones.

Let G_N be the SNR network of Definition 2.0.4 with parameters β and γ. Denote by $\mathrm{Comp}(G_N)$ the set of (connected) components of G_N. For any $\mathscr{C} \in \mathrm{Comp}(G_N)$ we write the graph as $(V(\mathscr{C}), E(\mathscr{C}))$ and denote by $|\mathscr{C}| := |V(\mathscr{C})|$ the number of vertices in $\mathscr{C}$. Moreover, we let $\mathscr{C}(i)$ denote the component containing vertex i. Throughout this section, we will use the notation

$$K_\gamma := N^{\frac{1-2\gamma}{2-2\gamma}} \log N,$$

and call a component $\mathscr{C} \in \mathrm{Comp}(G_N)$ *big* if $\mathscr{C} = \mathscr{C}(i)$ for some $i \leq K_\gamma$. Otherwise, the component is called *small*. Moreover, we define the collection of all vertices lying in big

components as

$$V_{\text{big}} := \bigcup_{i \le K_\gamma} V(\mathscr{C}(i)).$$

The first proposition is a standard result on the (componentwise) diameter.

Proposition 3.2.1. *For G_N with $\beta + 2\gamma < 1$, we have that*

$$\operatorname{diam}(G_N) := \sup_{\mathscr{C} \in \operatorname{Comp}(G_N)} \operatorname{diam}(\mathscr{C}) = O_{\mathbb{P}}(\log N).$$

As we will see later on, for the classical voter model, the invariant measure of the associated random walk is normalized by $\sum_{z \in \mathscr{C}(k)} \mathrm{d}(z)^{1-\theta}$, so that in the following we collect various bounds on $\sum_{z \in \mathscr{C}(k)} \mathrm{d}(z)$.

Proposition 3.2.2. *For G_N with $\beta + 2\gamma < 1$, with high probability,*

(a) "Big" components have size of comparable order to the degree of the vertex for which we call them big

$$\max_{k \le K_\gamma} \frac{\sum_{z \in \mathscr{C}(k)} \mathrm{d}(z)}{(N/k)^\gamma} \le \log N.$$

(b) "Small" components have a maximal polynomial size

$$\max_{i \notin V_{\text{big}}} \sum_{v \in \mathscr{C}(i)} \mathrm{d}(v) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right).$$

For the largest component this result is not optimal as we lose a log factor, see also [Janson, 2008, Theorem 1.1], but this result does not cover the other components.

As a next result, we need that the large degrees $\mathrm{d}(i)$ are well approximated by their means. Also, we need to know that for each of the vertices with large degree, a positive proportion of its neighbours has degree 1. One of the challenges in the proof is that we need these bounds uniformly over all big components.

Proposition 3.2.3. *For G_N with $\beta + 2\gamma < 1$ the following statements hold:*

(a)

$$\min_{k \le K_\gamma} \frac{\mathrm{d}(k)}{(N/k)^\gamma} = \Omega_{\mathbb{P}}(1), \qquad \max_{k \le K_\gamma} \frac{\mathrm{d}(k)}{(N/k)^\gamma} = O_{\mathbb{P}}(1).$$

(b) For any $k \in [N]$, let L_k be the number of neighbours of k of degree 1, then we have

$$\min_{k \le K_\gamma} \frac{|L_k|}{\mathrm{d}(k)} = \Omega_{\mathbb{P}}(1).$$

Definition 3.2.4. For G_N and any component $\mathscr{C} \in \mathrm{Comp}(G_N)$, we define the set of *branches* $\mathcal{B}(\mathscr{C})$ of $\mathscr{C}$ as the components created by deleting the vertex of minimal index, that is the set of connected components of the subgraph of $\mathscr{C}$ induced by the vertex set $V(\mathscr{C}) \setminus \{i\}$, where $i = \min V(\mathscr{C})$.

We will use this definition specifically in the context when $\mathscr{C}$ is a tree, so that this terminology makes sense. The next lemma states that big components are trees and have branches that are small (at least when compared to the largest components of order N^γ).

Lemma 3.2.5. *For G_N with $\beta + 2\gamma < 1$, with high probability every* big *component is a tree. On this event we have that*

$$\max_{k \le K_\gamma} \max_{B \in \mathcal{B}(\mathscr{C}(k))} \sum_{v \in B} \mathrm{d}(v) = \max_{k \le K_\gamma} \max_{B \in \mathcal{B}(\mathscr{C}(k))} (2|B| - 1) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right).$$

The following claim for the empirical moment of the degree distribution of $\mathscr{C}(1)$ will allow us to demonstrate a lower bound on this component for certain parameters of both the classical and discursive models.

Lemma 3.2.6. *For G_N with $\beta + 2\gamma < 1$, then for any $\eta \ge 1$*

$$\sum_{v \in \mathscr{C}(1)} \mathrm{d}(v)^\eta = \Theta_{\mathbb{P}}^{\log N}\left(N^{\gamma\eta}\right).$$

Two of our lower bounds require the existence of a 'double star' component together with a suitable bound on the empirical moment.

Proposition 3.2.7. *For G_N with $\beta + 2\gamma < 1$ there exists with high probability a tree component containing two adjacent vertices $x, y \in K_\gamma$ such that*

$$\mathrm{d}(x) \text{ and } \mathrm{d}(y) \text{ are } \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$$

and further for any $\eta \ge 1$

$$\sum_{v \in \mathscr{C}(x)} \mathrm{d}(v)^\eta = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma\eta}{2-2\gamma}}\right).$$

The final proposition of this section states that we can always find a "long double star" in G_N, i.e. two vertices with degree of order at least $N^{\gamma/(2-2\gamma)}$ that are connected via a short path with two intermediate vertices of degree 2 each. The path having length at least 3 is important for the discursive voter model dynamic.

Proposition 3.2.8. *With high probability any G_N with $\beta + 2\gamma < 1$ contains a path $\mathcal{P} = (v_1, v_2, v_3, v_4)$ such that:*

(a) $\mathrm{d}(v_2) = \mathrm{d}(v_3) = 2$.

(b) $\{v_1, v_4\} \subset [K_\gamma]$ *(and hence the component is a tree)*

(c) $\mathrm{d}(v_1), \mathrm{d}(v_4) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$.

In the remaining part of this section, we will prove these results. An essential tool will be a coupling with a branching process that we set up in Section 3.1. Then in Section 3.2 we will prove the structural results stated above.

Proofs for the simple Norros-Reittu network

In this section we will prove the statements made at the beginning of Section 3.2 using the coupling with a branching process as outlined in Section 3.1. A standard strategy will be to use that the SNR model can obtained from the MNR model by collapsing multi edges and then an upper bound on the degrees in the MNR model can be shown in terms of the Galton-Watson trees as described in Proposition 3.1.2.

Proof of Proposition 3.2.1. By the construction of Proposition 3.1.2, for an upper bound it suffices to bound the diameter in each component of $\mathcal{T}^1, \ldots, \mathcal{T}^n$ as extra edges are only created within components and so only make the diameters shorter, see also Remark 3.1.3.

Recall that in $\mathcal{T}^i$ the root has a $\mathrm{Pois}(w(i))$-distributed number of offspring, while the offspring distribution for any other vertex has the same distribution as $D \sim \mathrm{Pois}\,(W_N^*)$, where the offspring mean satisfies

$$\lim_{N\to\infty} \mathbb{E}\,(D) = \frac{\beta}{1-2\gamma} < 1. \tag{3.3}$$

Let $(Z_k)_{k\geq 0}$ be a Galton-Watson tree with offspring distribution D and $Z_0 = 1$. Then, for any $\rho \in \left(\frac{\beta}{1-2\gamma}, 1\right)$ and N large enough we have, by Markov's inequality,

$$\mathbb{P}(Z_k \neq 0) \leq \mathbb{E}(Z_k) \leq \rho^k.$$

By construction, it order to show the required bound on the diameter it suffices to bound the maximal depth of Y independent Galton-Watson trees $\mathcal{T}_i^*, i = 1, \dots, Y$, with offspring distribution D and where Y is Poisson-distributed random variable with parameter

$$w([N]) = \sum_{i,j \in [N]} \beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \sim \frac{\beta N}{(1-\gamma)^2}, \tag{3.4}$$

where here and in the following, we write $w(A) = \sum_{i \in [A]} w(i)$ for any $A \subset [N]$.

Since with high probability $Y \leq 2w([N]) \leq KN$ for a suitable constant $K \in \mathbb{N}$, we can get the required bound by noting that for any $C > 0$,

$$\begin{aligned}\mathbb{P}\Big(\max_{i=1,\dots,KN} \operatorname{diam}(\mathcal{T}_i^*) \geq C \log N\Big) &\leq \sum_{i=1}^{KN} \mathbb{P}(\operatorname{diam}(\mathcal{T}_1^*) \geq C \log N) \\ &\leq KN\, \mathbb{P}(Z_{\lfloor C \log N \rfloor} \neq 0) \leq KN \rho^{C \log N - 1},\end{aligned}$$

which converges to 0 if we choose C large enough such that $C \log \rho < 1$. □

In the following, we will develop a stochastic upper bound on the sizes of the trees $\mathcal{T}^i$ in Proposition 3.1.2 that no longer depends on N. Throughout we will write $X \preceq Y$ if the random variable Y stochastically dominates the random variable X.

Lemma 3.2.9. *For any $\alpha > 1$, $\gamma < \frac{1}{2}$ and N sufficiently large, we have*

$$W_N^* \preceq \alpha W^*,$$

where W^* is the weak limit of $(W_N^*)_N$ with density

$$\frac{\mathbb{P}(W^* \in \mathrm{d}x)}{\mathrm{d}x} = \mathbb{1}_{x > \frac{\beta}{1-\gamma}} \frac{1-\gamma}{\gamma} \left(\frac{\beta}{1-\gamma}\right)^{\frac{1}{\gamma}-1} x^{-\frac{1}{\gamma}}. \tag{3.5}$$

Proof. Note that the MNR weights satisfy for each $i \in [N]$,

$$w(i) = \beta N^{2\gamma-1} i^{-\gamma} \sum_{j=1}^{N} j^{-\gamma} \leq \frac{\beta}{1-\gamma} \left(\frac{N}{i}\right)^{\gamma} =: \lambda(i). \tag{3.6}$$

Moreover, by the definition of the distribution of the marks, we have that

$$\mathbb{P}(M \leq k) = \sum_{\ell=1}^{k} \frac{\ell^{-\gamma}}{\sum_{i=1}^{N} i^{-\gamma}} \leq \frac{1}{1-\gamma} \frac{k^{1-\gamma}}{\sum_{i=1}^{N} i^{-\gamma}}.$$

Now we can consider the tail of distribution function of $W_N^* \stackrel{\mathrm{d}}{=} w(M)$ and estimate for any $x \geq 0$,

$$\mathbb{P}\left(W_N^* \geq x\right) = \mathbb{P}\left(w(M) \leq x\right) \leq \mathbb{P}(\lambda(M) \leq x) = \mathbb{P}\left(M \leq \lambda^{-1}(x)\right) \leq \frac{\left(\lambda^{-1}(x)\right)^{1-\gamma}}{(1-\gamma)\sum_{i=1}^{N} i^{-\gamma}},$$

where we write $\lambda(x) := \frac{\beta}{1-\gamma} N^{\gamma} x^{-\gamma}$. Furthermore, we can compare this expression to

$$\mathbb{P}(\alpha W^* \geq x) = \left(\frac{1-\gamma}{\beta}\frac{x}{\alpha}\right)^{1-1/\gamma} = \alpha^{1/\gamma-1}\left(\frac{\lambda^{-1}(x)}{N}\right)^{1-\gamma}$$

Therefore we can conclude that

$$\inf_{x\geq 0} \frac{\mathbb{P}(\alpha W^* \geq x)}{\mathbb{P}\left(W_N^* \geq x\right)} \geq \alpha^{1/\gamma-1} \frac{(1-\gamma)\sum_{i=1}^{N} i^{-\gamma}}{N^{1-\gamma}} \to \alpha^{1/\gamma-1} > 1,$$

which gives the claimed stochastic domination. □

Proposition 3.2.10. *If $\alpha \in (1, \frac{1-2\gamma}{\beta})$, then $D_\alpha \sim \mathrm{Pois}\,(\alpha W^*)$ satisfies*

$$p_k := \mathbb{P}(D_\alpha = k) = \Theta(k^{-1/\gamma}).$$

In particular, if T is a Galton-Watson tree with offspring distribution $(p_k)_{k\geq 0}$, then the total size $|T|$ satisfies

$$\mathbb{P}(|T| = k) = \Theta\big(k^{-1/\gamma}\big).$$

Proof. The second statement follows from the first one by [Jonsson and Stefánsson, 2011, Thm. 4.1]. Note that the slightly weaker claim $\mathbb{P}(|T| \geq k) = \Theta\big(k^{1-1/\gamma}\big)$ follows from [Bagley, 1982, Theorem 2] with more general assumption on the degree, but we refer to [Jonsson and Stefánsson, 2011] here because we are really interested in the exact power law distribution.

To prove the first statement, we use that by (3.5) we know that αW^* has density f where $\mathrm{f}(x) = Cx^{-\frac{1}{\gamma}}\mathbb{1}_{x>\frac{\alpha\beta}{1-\gamma}}$, for some $C > 0$ that makes this a probability measure. Therefore,

$$p_{k-1} = \frac{C}{(k-1)!}\int_{\alpha\beta/(1-\gamma)}^{\infty} x^{k-1-1/\gamma}e^{-x}\mathrm{d}x = C\frac{\Gamma(k-1/\gamma) - E_k}{\Gamma(k)},$$

where the error term E_k is defined as

$$E_k = \int_0^{\alpha\beta/(1-\gamma)} x^{k-1-1/\gamma} e^{-x} \mathrm{d}x \leq \left(\frac{\alpha\beta}{1-\gamma}\right)^{k-1/\gamma} \leq 1,$$

and where the bound holds for $k \geq \frac{1}{\gamma} + 1$, using the assumption that α is small – hence $E_k = \Theta(1)$. Now we use $\Gamma(k) = \Theta\left(k^{k-1/2}e^{-k}\right)$ to rearrange

$$p_{k-1} = \Theta\left(\frac{(k-1/\gamma)^{k-1/\gamma-1/2}e^{1/\gamma-k}}{k^{k-1/2}e^{-k}}\right) = \Theta\left(\left(k - \frac{1}{\gamma}\right)^{-1/\gamma}\right)$$

by using the classical limit $\left(1 - \frac{1}{k\gamma}\right)^{k-1/2} \to e^{-1/\gamma}$. □

The statement of *Proposition 3.2.2 (a)* follows immediately from the following lemma, but we will also need the upper bound on the unthinned Galton-Watson forest.

Lemma 3.2.11. *For the Galton-Watson forest $(\mathcal{T}^1, \ldots, \mathcal{T}^N)$ as defined in the paragraph before Proposition 3.1.2, we have that with high probability*

$$\max_{k \leq K_\gamma} \frac{\sum_{z \in \mathcal{T}^k} \mathrm{d}(z)}{(N/k)^\gamma} \leq \log N.$$

Proof of Lemma 3.2.11. First, we consider the degrees of the roots, where we write $\varnothing_k$ for the root of the tree $\mathcal{T}^k$. Then, since $\mathrm{d}(\varnothing_k) \sim \mathrm{Pois}(w(k))$, standard large deviation estimates for Poisson distributions give a $C_1 > 0$ such that the event

$$E_1 := \left\{ \max_{k \leq K_\gamma} \frac{\mathrm{d}(\varnothing_k)}{(N/k)^\gamma} \leq C_1 \right\},$$

satisfies $\lim_{N\to\infty} \mathbb{P}(E_1) = 1$. Then, in distribution, each $\mathcal{T}^k$ consists of the root $\varnothing_k$ with $\mathrm{d}(\varnothing_K)$ edges to which we attach independent Galton-Watson trees, where the number of offspring has the same distribution as $\mathrm{Pois}(W_N^*)$. In particular, by Lemma 3.2.9, we can dominate the size of $\mathcal{T}^k$ by the size of $\mathcal{T}^{k,\alpha}$, a tree where the root $\varnothing_k$ has $\mathrm{d}(\varnothing_k)$ children, which are each connected to an independent Galton-Watson tree with offspring distribution $D_\alpha \sim \mathrm{Pois}(\alpha W^*)$, for some $\alpha \in (1, \frac{1-2\gamma}{\beta})$.

Now, if $T_1, \ldots, T_n$ denote independent copies of D_α-Galton-Watson trees, then by Proposition 3.2.10 the total sizes $|T_i|$ of these trees satisfy $\mathbb{P}(|T_i| = k) = \Theta(k^{-\frac{1}{\gamma}})$, so that

they are subexponential, see [Embrechts et al., 2013], in the sense that

$$\lim_{n\to\infty}\sup_{x\geq\gamma n}\left|1-\frac{\mathbb{P}\left(\sum_{k=1}^{n}\left(|T_k|-\mathbb{E}(|T_k|)\right)>x\right)}{n\mathbb{P}\left(|T_1|>x\right)}\right|=0.$$

In particular, we have that for any $\epsilon>0$, for all d sufficiently large and for $x\geq(\gamma+\mathbb{E}(|T_1|))d$

$$\mathbb{P}\Big(\sum_{k=1}^{d}|T_k|>x\Big)\leq(1+\epsilon)d\,\mathbb{P}(|T_1|>x-d\,\mathbb{E}(|T_1|)). \tag{3.7}$$

Since $\mathbb{E}(D_\alpha)=\frac{\alpha\beta}{1-2\gamma}$, we have that

$$\mathbb{E}(|T_1|)=\sum_{k=0}^{\infty}\Big(\frac{\alpha\beta}{1-2\gamma}\Big)^k=\frac{1}{1-\frac{\alpha\beta}{1-2\gamma}}.$$

Therefore, if we define $C_2=1+C_1\frac{\alpha\beta}{1-2\gamma}$, we have for any $k\leq K_\gamma$,

$$\begin{aligned}
\mathbb{P}\Big(|\mathcal{T}^k|>\Big(\frac{N}{k}\Big)^\gamma\log N;E_1\Big)&\leq\mathbb{P}\Big(|\mathcal{T}^{k,\alpha}|>\Big(\frac{N}{k}\Big)^\gamma\log N;E_1\Big)\\
&\leq\mathbb{P}\left(1+\sum_{i=1}^{\lfloor C_1(\frac{N}{k})^\gamma\rfloor}|T_i|>\Big(\frac{N}{k}\Big)^\gamma\log N\right)\\
&\leq(1+\epsilon)C_1\Big(\frac{N}{k}\Big)^\gamma\mathbb{P}\left(|T_1|>\Big(\frac{N}{k}\Big)^\gamma\log N-C_2\Big(\frac{N}{k}\Big)^\gamma\right)\\
&\leq(1+2\epsilon)C_1\Big(\frac{N}{k}\Big)^\gamma\mathbb{P}\left(|T_1|>\Big(\frac{N}{k}\Big)^\gamma\log N\right),
\end{aligned}$$

where the last inequality holds for N sufficiently large because we know $|T_1|$ has a power law tail. Hence, by a union bound

$$\begin{aligned}
&\mathbb{P}\Big(\exists k\leq K_\gamma:|\mathcal{T}^k|>\Big(\frac{N}{k}\Big)^\gamma\log N\Big)\\
&\qquad\leq(1+2\epsilon)C_1N^\gamma\sum_{k\leq K_\gamma}k^{-\gamma}\,\mathbb{P}\left(|T_1|>\Big(\frac{N}{k}\Big)^\gamma\log N\right),
\end{aligned}$$

which for N large enough we can bound for some suitable constant $C_3 > 0$ by

$$\begin{aligned}
&(1+2\epsilon)C_3 N^{\gamma} \sum_{k \le K_\gamma} k^{-\gamma} \left(\left(\frac{N}{k} \right)^{\gamma} \log N \right)^{\frac{\gamma-1}{\gamma}} \\
&= (1+2\epsilon)C_3 N^{2\gamma-1} \log^{\frac{\gamma-1}{\gamma}} N \sum_{k \le K_\gamma} k^{1-2\gamma} \\
&\le \frac{(1+2\epsilon)C_3}{2-2\gamma} \log^{3-\frac{1}{\gamma}-2\gamma} N = o(1),
\end{aligned}$$

where we used $K_\gamma = N^{\frac{1-2\gamma}{2-2\gamma}} \log N$. From this bound on the number of vertices, we can immediately deduce the claimed bound for the sum of the degrees, since

$$\sum_{v \in \mathcal{T}^k} \mathrm{d}(v) = 2\left|\mathcal{T}^k\right| - 2,$$

as each $\mathcal{T}^k$ is a tree. □

Proposition 3.2.12. *On G_N with $\beta + 2\gamma < 1$ we have the following uniform bound on the degrees of vertices with larger index*

$$\max_{k > K_\gamma} \mathrm{d}(k) = O_{\mathbb{P}}\left(N^{\frac{\gamma}{2-2\gamma}} \right).$$

Proof. As before we can use that the degrees are stochastically dominated by the degrees for the MNR model, where each $\mathrm{d}(k) \sim \mathrm{Pois}(w(k))$. The Chernoff bound then gives that

$$\begin{aligned}
\log \mathbb{P}\left(d(k) \ge N^{\frac{\gamma}{2-2\gamma}} \right) &\le \log \mathbb{P}\left(\mathrm{Pois}(w(k)) \ge N^{\frac{\gamma}{2-2\gamma}} \right) \\
&\le \log \mathbb{E}\left(e^{\mathrm{Pois}(w(k))} \right) - N^{\frac{\gamma}{2-2\gamma}} \\
&= w(k)(e-1) - N^{\frac{\gamma}{2-2\gamma}}.
\end{aligned}$$

where by a slight abuse of notation we also write $\mathrm{Pois}(w(k))$ for a Poisson random variable with parameter $w(k)$. Hence, by a union bound and using that $w(k)$ is decreasing, we

obtain

$$\begin{aligned}\mathbb{P}\left(\exists k > K_\gamma : d(k) > N^{\frac{\gamma}{2-2\gamma}}\right) &\leq e^{-N^{\frac{\gamma}{2-2\gamma}}} \sum_{k=K_\gamma+1}^{N} e^{w(k)(e-1)} \\ &\leq (N - K_\gamma + 1) e^{w(K_\gamma)(e-1) - N^{\frac{\gamma}{2-2\gamma}}} \\ &\leq N \exp\left(N^{\frac{\gamma}{2-2\gamma}} \left(-1 + \frac{\beta(e-1)}{1-\gamma} \frac{1}{\log^\gamma N}\right)\right) = o(1),\end{aligned}$$

where in the last step we used that $w(k) \leq \frac{\beta}{1-\gamma} \left(\frac{N}{k}\right)^\gamma$, see also (3.6). □

Proof of Proposition 3.2.2 (b). As before it suffices to bound the degrees in the MNR graph. By Remark 3.1.4, the tree construction yields the stochastic upper bound

$$\max_{k \notin V_{\text{big}}} \sum_{v \in \mathscr{C}(k)} \mathrm{d}(v) \preceq \max_{k=K_\gamma+1,\dots,N} \sum_{v \in \mathcal{T}_k^k} \mathrm{d}(v) \leq \max_{k=K_\gamma+1,\dots,N} 2|\mathcal{T}_k^k|,$$

where $\mathcal{T}_k^k$ are independent Galton-Watson trees with the following law: the root of $\mathcal{T}_k^k$ has a $\mathrm{Pois}(w(k))$ number of offspring and all other offspring are independent and have a $\mathrm{Pois}(W_{N,k}^*)$ distribution.

For any thinning level $z \in [N]$, we recall that $W_{N,z}^*$ is defined as

$$W_{N,z}^* \overset{(\mathrm{d})}{=} w(M) \mathbb{1}_{M>z}, \tag{3.8}$$

where M is the usual mark distribution (which chooses $i \in [N]$ with probability proportional to $i^{-\gamma}$).

The same argument as in the proof of Proposition 3.2.12 also shows that there exists a constant $C_1 > 0$ such that if $\circledcirc_k$ denotes the root of $\mathcal{T}_k^k$, then we have that the event

$$E_1 := \left\{ \max_{k=K_\gamma+1,\dots,N} \mathrm{d}(\circledcirc_k) \leq C_1 N^{\frac{\gamma}{2-2\gamma}} \right\} \tag{3.9}$$

satisfies $\mathbb{P}(E_1) \to 1$ as $N \to \infty$.

To bound the size of the trees $\mathcal{T}_k^k$, we use the standard connection to random walks, see e.g. [van der Hofstad, 2016, Section 3.3], where we consecutively record the number of offspring of each individual in the branching process.

Define

$$R := \left\lfloor C_1 N^{\frac{\gamma}{2-2\gamma}} \right\rfloor,$$

for the same constant C_1 as in (3.9).

Then, define a random walk $(S_n)_{n\geq 0}$, where we set $S_0 = 0, S_1 = R$ and for $i \geq 1$ suppose that the increments $S_{i+1} - S_i$ are independent and with the same distribution as $D-1$, where $D \sim \text{Pois}(W^*_{N,z})$. The random walk connection then yields that for any $z \in \{K_\gamma + 1, \dots, N\}$ and any $L \geq 0$

$$\mathbb{P}(|\mathcal{T}_z^z| > L; E_1) \leq \mathbb{P}(S_{L+1} \geq 0).$$

Now for large N we define

$$L := \frac{2R}{1 - \mathbb{E}(W_N^*)},$$

which is well-defined as $\mathbb{E}(W_N^*) \to \mathbb{E}(W^*) < 1$. Moreover, we define $(X_i^{(z)})_{i\geq 1}$ as a sequence of i.i.d. random variables with the same distribution as $D - \mathbb{E}(W^*_{N,z})$, where $D \sim \text{Pois}(W^*_{N,z})$. Then we can estimate using $\mathbb{E}(W^*_{N,z}) \leq \mathbb{E}(W_{N,z})$ that

$$\begin{aligned}
\mathbb{P}\left(|\mathcal{T}_z^z| > L; E_1\right) &\leq \mathbb{P}\bigg(R + \sum_{i=1}^{L}(S_{i+1} - S_i) > L\bigg) \\
&\leq \mathbb{P}\bigg(\sum_{i=1}^{L} X_i^{(z)} \geq L\Big(1 - \mathbb{E}(W^*_{N,z})\Big) - R\bigg) \\
&\leq \mathbb{P}\bigg(\sum_{i=1}^{L} X_i^{(z)} \geq R\bigg).
\end{aligned}$$

Then, by Markov's inequality for any $r > 2 \vee \left(\frac{1}{\gamma} - 1\right)$ and N sufficiently large, we can deduce that

$$\mathbb{P}\left(|\mathcal{T}_z^z| > L; E_1\right) \leq \frac{\mathbb{E}|\sum_{i=1}^{L} X_i^{(z)}|^r}{R^r} \leq C_2 \frac{L^{\frac{r}{2}} w(z)^{\frac{r}{2}\left(3-\frac{1}{\gamma}\right)^+} + L w(z)^{r+1-\frac{1}{\gamma}}}{R^r}, \tag{3.10}$$

where $C_2 > 0$ is a suitable constant, coming out of the estimate on the fractional moment, which we defer to Lemma 3.2.13.

For the remainder of the proof, we will need to fix an even larger r and assume that

$$r > \frac{4 - 4\gamma}{\gamma \wedge (1 - 2\gamma)}.$$

By a union bound combined with the bound in (3.10) and the definitions of L and

R, we find that there exists a $C_3 > 0$ such that

$$\begin{aligned}\mathbb{P}\Big(\max_{z\in\{K_\gamma+1,\dots,N\}} |\mathcal{T}_z^z| > L; E_1\Big) &\leq \sum_{z=K_\gamma+1}^{N} \mathbb{P}\left(|\mathcal{T}_z^z| > L; E_1\right) \\ &\leq C_3 \sum_{z=K_\gamma+1}^{N} \frac{N^{\frac{r}{2}(3\gamma-1)^+ - \frac{r}{2}\frac{\gamma}{2-2\gamma}}}{z^{\frac{r}{2}\left(3-\frac{1}{\gamma}\right)^+ \log^{\frac{r}{2}} N}} + \frac{N^{\gamma r+\gamma-1+(1-r)\frac{\gamma}{2-2\gamma}}}{z^{\gamma r+\gamma-1}\log^{r-1} N}\end{aligned} \tag{3.11}$$

and we require that this sum tends to 0. For the first term, observe

$$\frac{N^{\frac{r}{2}(3\gamma-1)^+ - \frac{r}{2}\frac{\gamma}{2-2\gamma}}}{z^{\frac{r}{2}\left(3-\frac{1}{\gamma}\right)^+}} \leq \frac{N^{\frac{r}{2}(3\gamma-1)^+ - \frac{r}{2}\frac{\gamma}{2-2\gamma}}}{N^{\frac{1-2\gamma}{2-2\gamma}\frac{r}{2}\left(3-\frac{1}{\gamma}\right)^+}} = \left(N^{\frac{(3\gamma-1)^+-\gamma}{2-2\gamma}}\right)^{\frac{r}{2}},$$

and we note that the exponent of N in this expression is less than -1 by our choice of r and since

$$(3\gamma-1)^+ - \gamma = \begin{cases} 2\gamma-1, & \text{if } \gamma > \frac{1}{3}, \\ -\gamma, & \text{if } \gamma \leq \frac{1}{3}. \end{cases}$$

In particular, the first term in (3.11) converges to 0.

For the second term, again by our choice of r, we have that $r > \frac{2-\gamma}{\gamma}$ so that we can deduce that

$$\sum_{z=K_\gamma+1}^{N} \frac{1}{z^{\gamma r+\gamma-1}} = O\left(N^{\frac{1-2\gamma}{2-2\gamma}(2-\gamma r-\gamma)}\right).$$

In particular, we have that

$$\begin{aligned}\frac{N^{\gamma r+\gamma-1+(1-r)\frac{\gamma}{2-2\gamma}}}{\log^{r-1} N} \sum_{z>K_\gamma} \frac{1}{z^{\gamma r+\gamma-1}} &= O\left(\frac{N^{\gamma r+\gamma-1+(1-r)\frac{\gamma}{2-2\gamma}}}{\log^{r-1} N} N^{\frac{1-2\gamma}{2-2\gamma}(2-\gamma r-\gamma)}\right) \\ &= O\left(\log^{1-r} N\right) = o(1),\end{aligned}$$

which shows that also the second term in (3.11) tends to 0.

Thus, we can conclude from (3.11) that with high probability on the event E_1 every tree has size at most L. Recalling that E_1 occurs also with high probability and the asymptotics of L then gives the required bound. □

The following lemma provides the moment estimate that was required in the proof of Proposition 3.2.2 (b) above.

Lemma 3.2.13. *For $L \in \mathbb{N}, z \in [N]$, suppose $X_i^{(z)}, i \leq L$ are independent random variables with the same distribution as $D - \mathbb{E}(W_{N,z}^*)$, where $D \sim \mathrm{Pois}(W_{N,z}^*)$ and where*

$W^*_{N,z}$ is defined in (3.8). *Then, for any $L, N \in \mathbb{N}$, $r > 2 \vee \left(\frac{1}{\gamma} - 1\right)$ and $z > 1$ we have*

$$\mathbb{E}\left(\Big|\sum_{i=1}^{L} X_i^{(z)}\Big|^r\right) \leq C\Big(L^{\frac{r}{2}} w(z)^{\frac{r}{2}\left(3-\frac{1}{\gamma}\right)^+} + L w(z)^{r+1-\frac{1}{\gamma}}\Big),$$

where $C > 0$ is a constant depending only on r and γ and $w(z)$ is defined in (3.1).

Proof. These calculations use a similar strategy to the proof of [Janson, 2008, Theorem 1.1]. We write $X^{(z)}$ for a random variable with the same distribution as $X_i^{(z)}$. We start by estimating the second and the rth moment of $X^{(z)}$. First, note that as

$$\mathbb{E}\left(W^*_{N,z}\right) \leq \mathbb{E}\left(W^*_N\right) \to \frac{\beta}{1-2\gamma},$$

we deduce for N sufficiently large

$$\begin{aligned}\mathbb{E}\big((X^{(z)})^2\big) &= \mathrm{Var}\left(\mathrm{Pois}\left(X^{(z)}\right)\right) = \mathrm{Var}\left(\mathbb{E}\left(X^{(z)}|W^*_{N,z}\right)\right) + \mathbb{E}\left(\mathrm{Var}\left(X^{(z)}|W^*_{N,z}\right)\right)\\ &= \mathrm{Var}\left(W^*_{N,z}\right) + \mathbb{E}\left(W^*_{N,z}\right) \leq \mathbb{E}\left(\left(W^*_{N,z}\right)^2\right) + \frac{\beta}{1-2\gamma}.\end{aligned}$$

Using Lemma 3.2.9, we find a constant $C_1 > 0$ independent of N for any $\alpha > 1$ such that

$$\begin{aligned}\mathbb{E}\left((W^*_{N,z})^2\right) &= \int_0^\infty 2x\,\mathbb{P}(W^*_{N,z} > x)\,\mathrm{d}x = \left(\frac{\beta}{1-\gamma}\right)^2 + \int_{\frac{\beta}{1-\gamma}}^{w(z)} 2x\,\mathbb{P}(W^*_N > x)\,\mathrm{d}x\\ &\leq \left(\frac{\beta}{1-\gamma}\right)^2 + \int_{\frac{\beta}{1-\gamma}}^{w(z)} 2x\,\mathbb{P}(\alpha W^* > x)\,\mathrm{d}x \leq C_1 \int_{\frac{\beta}{1-\gamma}}^{w(z)} x^{2-\frac{1}{\gamma}}\,\mathrm{d}x\\ &\leq C_1 w(z)^{\left(3-\frac{1}{\gamma}\right)^+},\end{aligned}$$

for N sufficiently large, where we used the explicit density of W^* identified in (3.5). We now have to estimate $\mathbb{E}\big((X^{(z)})^r\big)$ for r as above. We claim that

$$\sup_{\lambda \geq \frac{\beta}{1-\gamma}} \frac{\mathbb{E}(\mathrm{Pois}(\lambda)^r)}{\lambda^r} < \infty. \tag{3.12}$$

Indeed, since the Poisson distribution has an exponential moment, we know $\lambda \mapsto$

$\mathbb{E}(\text{Pois}(\lambda)^r)/\lambda^r$ is finite and continuous on $\left[\frac{\beta}{1-\gamma}, \infty\right)$. Further

$$\left\| \frac{\text{Pois}(\lambda)}{\lambda} \right\|_r \leq \frac{\lambda + \|\text{Pois}(\lambda) - \lambda\|_r}{\lambda} = 1 + \frac{\left\| \frac{\text{Pois}(\lambda)-\lambda}{\sqrt{\lambda}} \right\|_r}{\sqrt{\lambda}} \to 1$$

as $\lambda \to \infty$, by the central limit theorem. This proves the claim (3.12). Hence we can say because $X^{(z)} \preceq \text{Pois}\left(W^*_{N,z}\right)$ we have some $C_2 > 0$ such that

$$\mathbb{E}\left(\left(X^{(z)}\right)^r\right) \leq C_2 \mathbb{E}((W^*_{N,z})^r)$$

for N sufficiently large. We have, given $r > \frac{1}{\gamma} - 1$,

$$\begin{aligned}
\frac{\mathbb{E}((W^*_{N,z})^r)}{z\mathbb{P}(M=z)w(z)^r} = \frac{\sum_{j=z}^N \mathbb{P}(M=j)w(j)^r}{z\mathbb{P}(M=z)w(z)^r} &= \sum_{j=z}^N \frac{1}{z} \frac{j^{-\gamma} j^{-\gamma r}}{z^{-\gamma} z^{-\gamma r}} \\
&\leq z^{\gamma+\gamma r-1} \int_{z-1}^{\infty} j^{-\gamma-\gamma r} \mathrm{d}j = \frac{1}{\gamma+\gamma r-1} \left(1 - \frac{1}{z}\right)^{1-\gamma-\gamma r} \\
&\leq \frac{1}{\gamma+\gamma r-1} \left(1 - \frac{1}{N}\right)^{1-\gamma-\gamma r} = O(1),
\end{aligned}$$

as $N \to \infty$, noting that this bound holds uniformly over all $z > 1$. Thus there is some constant C_3 such that, for every $z > 1$,

$$\mathbb{E}\left(\left(X^{(z)}\right)^r\right) < C_3 z \mathbb{P}(M=z) w(z)^r.$$

Therefore by Rosenthal's inequality [Gut, 2013, Chapter 3, Theorem 9.1] we obtain for $r > 2 \vee (\frac{1}{\gamma} - 1)$

$$\begin{aligned}
\mathbb{E}\left(\left| \sum_{i=1}^L \left(X_i^{(z)}\right) \right|^r\right) &\leq C_3 L^{r/2} \big(\mathbb{E}\big(\big((X^{(z)})^2\big)\big)^{r/2} + C_4 L \mathbb{E}\big((X^{(z)})^r\big) \\
&\leq C_5 L^{r/2} \left(1 + w(z)^{\left(3-\frac{1}{\gamma}\right)^+}\right)^{r/2} + C_6 L z \mathbb{P}(M=z) w(z)^r \\
&\leq C_7 L^{r/2} w(z)^{\left(3-\frac{1}{\gamma}\right)^+ \frac{r}{2}} + C_8 L \frac{z^{1-\gamma} w(z)^r}{N^{1-\gamma}},
\end{aligned}$$

as claimed. □

Proof of Proposition 3.2.3. (a) Again we construct the SNR network G_N via the MNR network and use the tree construction of Proposition 3.1.2 for an upper bound. Recall

by a standard concentration argument, there is some universal constant $C > 0$ such that for λ large enough

$$\mathbb{P}\left(|\mathrm{Pois}\,(\lambda) - \lambda| \leq \frac{\lambda}{2}\right) \leq e^{-C\lambda}, \tag{3.13}$$

so that for the unthinned degrees in $\mathcal{T}^k$ we can immediately compare $d(k)$ to $(N/k)^\gamma$. The upper bound $d(k)$ in the MNR network then follows immediately, because thinning can only decrease the degree.

For a lower bound, it suffices to show that overall not too many vertices are being thinned in the big components. More precisely, define

$$Z^{\mathrm{thin}} := \sum_{k=1}^{K_\gamma} \sum_{v \in \mathcal{T}^k} \mathbb{1}_{\{v \text{ thinned}\}}.$$

We will show in the following $Z^{\mathrm{thin}} \leq (\log N)^3$ with high probability, which immediately implies the lower bound on the degres (as these are polynomially large) by the same concentration argument for the Poisson degrees as before.

From Lemma 3.2.11 we have with high probability

$$\max_{k \leq K_\gamma} \frac{|\mathcal{T}^k|}{(N/k)^\gamma} \leq \max_{k \leq K_\gamma} \frac{\sum_{z \in \mathcal{T}^k} \mathrm{d}(z)}{(N/k)^\gamma} \leq \log N.$$

Therefore, by summation and the definition of K_γ we have that the event

$$E_1 = \Big\{ \Big| \bigcup_{i \leq K_\gamma} V(\mathscr{C}(i)) \Big| = |V_{\mathrm{Big}}| \leq \frac{1}{1-\gamma} \sqrt{N} \log N \Big\},$$

satisfies $\mathbb{P}(E_1) \to 1$ as $N \to \infty$.

We can bound Z^{thin} by the double sum over vertices that have the same mark. Thus, if we write M, M' for two independent copies of the mark distribution, then by distinguishing the cases of root vertices and remaining vertices, we obtain

$$\begin{aligned} \mathbb{E}(Z^{\mathrm{thin}} \mathbb{1}_{E_1}) &\leq \frac{1}{1-\gamma} \sqrt{N} (\log N) \mathbb{P}\left(M \leq K_\gamma\right) + \frac{1}{(1-\gamma)^2} N (\log N)^2 \mathbb{P}\left(M = M'\right) \\ &= O\left(\log^{2-\gamma} N + \log^2 N\right), \end{aligned}$$

where we used that

$$\mathbb{P}(M = M') = \sum_{i=1}^{N} \left(\frac{i^{-\gamma}}{\sum_{j=1}^{N} j^{-\gamma}} \right)^2 = \Theta\left(\frac{N^{1-2\gamma}}{N^{2-2\gamma}} \right) = \Theta\left(\frac{1}{N} \right). \tag{3.14}$$

Hence, by Markov's inequality, we have that

$$\mathbb{P}(Z^{\text{thin}} \geq (\log N)^3) \leq \frac{\mathbb{E}(Z^{\text{thin}} \mathbb{1}_{E_1})}{(\log N)^3} + \mathbb{P}(E_1^c),$$

and the right hand side tends to zero as $N \to \infty$.

(b) Now each of the neighbours of the root vertices $\{\omega_k\}$ which was not thinned has offspring $D \sim \text{Pois}(W_N^*)$ and independently no children with probability

$$p_0 = \mathbb{P}(D = 0) = \mathbb{E}(e^{-W_N^*}) \geq e^{-\mathbb{E}(W_N^*)} \to e^{-\frac{\beta}{1-2\gamma}} > 0$$

by using Jensen's inequality. This gives the bound on $|L_k|$ by a binomial concentration argument. □

Proof of Lemma 3.2.5. By Lemma 3.2.11 we know that the event

$$E_1 := \left\{ |\mathscr{C}(k)| \leq \left(\frac{N}{k}\right)^{\gamma} \log N \text{ for all } k \leq K_\gamma \right\}, \tag{3.15}$$

satisfies $\mathbb{P}(E_1) \to 1$ as $N \to \infty$.

As observed in Remark 3.1.3, the thinning operation does not create cycles between components, nor does it create extra edges between the root vertex and one of its children.

We now bound the surplus of each component, which is defined as the number of edges more than edges of a tree on the same vertex set. Writing M and M' for two independent copies of the mark distribution, we get

$$\mathbb{E}(\text{surplus}(\mathscr{C}(k)); E_1) \leq \mathbb{P}(M = M') \left(\frac{N}{k}\right)^{2\gamma} \log^2 N. \tag{3.16}$$

Hence, we obtain for the total surplus in the big components, using (3.14)

$$\begin{aligned}
\sum_{i=1}^{\lfloor K_\gamma \rfloor} \mathbb{E}(\text{surplus}(\mathscr{C}(i)); E_1) &\leq \mathbb{P}(M = M') N^{2\gamma} \log^2 N \int_0^{K_\gamma} i^{-2\gamma} \mathrm{d}i \\
&= O^{\log N}\left(N^{2\gamma - 1} N^{\frac{(1-2\gamma)^2}{2-2\gamma}} \right) \\
&= O^{\log N}\left(N^{\frac{2\gamma-1}{2-2\gamma}} \right) = o(1).
\end{aligned}$$

Combining this with the fact that $\mathbb{P}(E_1) \to 1$, we obtain by Markov's inequality that the big components form a forest with high probability.

Now that we know each component is a tree, it makes sense to talk about *branches* of the root vertices. Again, we can stochastically upper bound the sizes of these branches in SNR by the ones in MNR, which are bounded by the (unthinned) branches in the forest $\mathcal{T}^1, \dots, \mathcal{T}^{\lfloor K_\gamma \rfloor}$. In the latter, each of the branches is an independent Pois(W_N^*)-GW tree. Note that the total number of these trees is bounded by $\sum_{k=1}^{\lfloor K_\gamma \rfloor} \mathrm{d}(\varnothing_k)$, where $\varnothing_k$ is the root of $\mathcal{T}^k$. By the same argument as in the proof of Lemma 3.2.11, we therefore have that there exists a constant $C_2 > 0$ such that the event

$$E_2 = \Big\{ \sum_{k=1}^{\lfloor K_\gamma \rfloor} \mathrm{d}(\varnothing_k) \le C_2 \sqrt{N} \log^{1-\gamma} N \Big\},$$

satisfies $\mathbb{P}(E_2) \to 1$, since $\sum_{i=1}^{\lfloor K_\gamma \rfloor} \left(\frac{N}{i}\right)^\gamma = O(\sqrt{N} \log^{1-\gamma} N)$. Let $(T_i)_{i \ge 1}$ be a sequence of i.i.d. Pois(αW^*)-GW trees, where $\alpha \in \left(1, \frac{1-2\gamma}{\beta}\right)$. Further, let $J = \left\lfloor C_2 \sqrt{N} \log N \right\rfloor$. Then, by the above argument and Lemma 3.2.9, we have that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{k \le K_\gamma} \max_{B \in \mathcal{B}(\mathscr{C}(k))} \sum_{v \in B} \mathrm{d}(v) \ge N^{\frac{\gamma}{2-2\gamma}} \log N; E_2 \Big) \\
&\qquad \le \mathbb{P}\left(\max_{i=1}^{J} |T_i| > N^{\frac{\gamma}{2-2\gamma}} \log N \right) \\
&\qquad \le \sum_{i=1}^{J} \mathbb{P}(|T_1| > N^{\frac{\gamma}{2-2\gamma}} \log N) \\
&\qquad = O\Big(J \big(N^{\frac{\gamma}{2-2\gamma}} \log N \big)^{-(\frac{1}{\gamma}-1)} \Big) = O\left(\log(N)^{2-\frac{1}{\gamma}} \right) = o(1),
\end{aligned}$$

where we used Proposition 3.2.10 in the final line. Since $\mathbb{P}(E_2^c) \to 0$, this completes the proof of the lemma. □

Proof of Lemma 3.2.6. Note that $\mathrm{d}(\varnothing_1)$, the degree of the root of $\mathcal{T}^1$, satisfies $\mathrm{d}(\varnothing_1) \sim$ Pois$(w(1))$ and $w(1) = \Theta(N^\gamma)$. Hence, we can immediately deduce by standard Poisson concentration that there are constants $c_1, C_1 > 0$ such that the event

$$E_1 := \{ c_1 N^\gamma \le \mathrm{d}(\varnothing_1) \le C_1 N^\gamma \},$$

holds with high probability.

For a lower bound, we note that for the SNR model

$$\sum_{v \in \mathscr{C}(1)} \mathrm{d}(v)^\eta \ge \mathrm{d}(1)^\eta.$$

However, on the event E_1 for $\mathcal{T}_1$ the expected number of repeated labels among the children of the root $\varnothing_1$ is of order

$$O^{\log N}\left(N^{\gamma}\mathbb{P}(M=1)+N^{2\gamma}\mathbb{P}(M=M')\right)=O^{\log N}\left(N^{2\gamma-1}\right)=o(1),$$

where as before we write M, M' for independent copies of the mark distribution. So with high probability we have that $\mathrm{d}(1)=\mathrm{d}(\varnothing_1)\geq c_1 N^{\gamma}$, which gives the required lower bound.

For the upper bound, by the same arguments used before we only have to bound the degrees in $\mathcal{T}^1$. Note that with the exception of the degree of the root, all other degrees have the same distribution as D, where $D\sim 1+\mathrm{Pois}\,(W_N^*)$. Note also that by Lemma 3.2.11 there exists a constant C_2 such that the event $E_2=\{|\mathcal{T}^1|\leq C_2 N^{\gamma}\log N\}$ occurs with high probability. Therefore, if we let $(D_i)_{i\geq 1}$ be i.i.d. random variables with the same distribution as D and set $J:=\lfloor C_2 N^{\gamma\log N}\rfloor$, then for any $\eta>1$.

$$\mathbb{P}\left(\sum_{v\in\mathcal{T}^1,v\neq\varnothing_1}\mathrm{d}(v)^{\eta}\geq(\log N)^3N^{2\gamma};E_2\right)\leq\mathbb{P}\left(\sum_{i=1}^{J}D_i^{\gamma}\geq(\log N)^3N^{\eta\gamma}\right).$$

We estimate the probability on the right by a first moment bound. Notice that

$$\mathbb{E}((W_N^*)^{\eta})=\sum_{i=1}^{N}\frac{w(i)^{\eta+1}}{\sum_{j=1}^{N}w(j)}=\Theta\begin{cases}N^{\gamma(\eta+1)-1} & \eta>\frac{1}{\gamma}-1,\\ \log N & \eta=\frac{1}{\gamma}-1,\\ 1 & \eta<\frac{1}{\gamma}-1.\end{cases}$$

Moreover, there exists a constant $C_3>0$ such that

$$\mathbb{E}\left((1+\mathrm{Pois}\,(W_N^*))^{\eta}\right)\leq 2^{\eta}\mathbb{E}\left(1\vee\mathrm{Pois}\,(W_N^*)^{\eta}\right)\leq C_3\mathbb{E}\left((W_N^*)^{\eta}\right),$$

and in particular we have that

$$\mathbb{E}\left(\sum_{i=1}^{J}D_i^{\eta}\right)=O\left((\log N)^2N^{(\gamma(\eta+1)-1)^+\gamma}\right)\leq O\left((\log N)^2N^{\eta\gamma}\right),$$

where we used that $\gamma<\frac{1}{2}$ in the last step. Hence, by Markov's inequality,

$$\mathbb{P}\left(\sum_{v\in\mathcal{T}^1,v\neq\varnothing_1}\mathrm{d}(v)^{\eta}\geq(\log N)^3N^{\eta\gamma};E_2\right)\leq O((\log N)^{-1}).$$

Since the event E_2 occurs with high probability and since we have by the first part that $\mathrm{d}(\circledcirc_1)^\eta \leq C_1^\eta N^{\eta\gamma}$ on the high probability event E_1, the upper bound in the statement of the lemma follows immediately. □

Proof of Proposition 3.2.7. Consider the index set

$$I := \left[N^{\frac{1-2\gamma}{2-2\gamma}}, N^{\frac{1-2\gamma}{2-2\gamma}} \log N\right] \cap \mathbb{N} \subset [K_\gamma].$$

We calculate the MNR weight as

$$w(I) \sim \sum_{k\in I} \frac{\beta}{1-\gamma}\left(\frac{N}{k}\right)^\gamma = \Theta\left(\sqrt{N}\log^{1-\gamma} N\right)$$

and so in for the MNR model the expected number of edges on the subgraph induced on I is

$$\frac{w(I)^2}{w([N])} = \Theta\left(\log^{2-2\gamma} N\right),$$

and so diverges to ∞. Since this number is Poisson distributed, it must then be nonzero with high probability. After collapsing any multi-edges to arrive at the SNR model it must still be nonzero with high probability.

We can take any such adjacent pair $(x, y) \in I^2$ to create a double star, which by Lemma 3.2.5 is a tree and by Proposition 3.2.3 (a) has

$$\mathrm{d}(x) \text{ and } \mathrm{d}(y) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right).$$

For the final claim of the Proposition we consider the empirical moment. The estimate on the moment can be proved in the same way as in the previous proof of Lemma 3.2.6, but instead we now have to control the ηth empirical moment of an i.i.d. sequence $D_i, i = 1, \ldots, \lfloor N^{\frac{\gamma}{2-2\gamma}} \log N \rfloor$ where $D_i \sim 1 + \mathrm{Pois}\left(W_N^*\right)$. Then, the result follows by analogous argument, combined with the fact that with high probability we do not see any thinning on this double star. □

Proof of Proposition 3.2.8. We first prove the statement for the multigraph MNR. Let V be the set of vertices with weight w (as defined in (3.1)) less than 1,

$$V := \{v \in [N] : w(v) < 1\}.$$

As in the previous proof, we consider $I := \left[N^{\frac{1-2\gamma}{2-2\gamma}}, N^{\frac{1-2\gamma}{2-2\gamma}} \log N\right] \cap \mathbb{N} \subset [K_\gamma]$. We

split both vertex sets into even and odd vertices as

$$V^{\text{even}} := V \cap (2\mathbb{N})\,; \qquad V^{\text{odd}} := V \cap (2\mathbb{N}+1)\,;$$

$$I^{\text{even}} := I \cap (2\mathbb{N})\,; \qquad I^{\text{odd}} := I \cap (2\mathbb{N}+1)\,.$$

Recall that $w(v) \le \frac{\beta}{1-\gamma}\left(\frac{N}{v}\right)^{\gamma}$ by (3.6), so any vertex v with $v > N\left(\frac{\beta}{1-\gamma}\right)^{1/\gamma}$ is in V. Since we also that have by assumption $\frac{\beta}{1-\gamma} < \frac{1-2\gamma}{1-\gamma} < 1$, we can conclude that $|V| = \Theta(N)$. Thus as $N \to \infty$,

$$\frac{w(V^{\text{even}})}{N} \sim \frac{w(V^{\text{odd}})}{N} \to \rho > 0,$$

where as before we write $w(A) = \sum_{i\in A} w(i)$ for any $A \subset [N]$. We also recall from (3.4) that $w([N]) \sim \frac{\beta N}{(1-\gamma)^2} = \Theta(N)$ and finally for the large degree sets, we get

$$w(I^{\text{even}}) \sim w(I^{\text{odd}}) \sim \frac{1}{2}\sum_{k\in I} \frac{\beta}{1-\gamma}\left(\frac{N}{k}\right)^{\gamma} \sim \frac{\beta}{2(1-\gamma)^2} N^{\gamma} K_{\gamma}^{1-\gamma}.$$

The number of edges from V^{even} to I^{even} in the MNR model is Poisson distributed with mean

$$\frac{w(V^{\text{even}})w(I^{\text{even}})}{w([N])} = \Theta\left(\frac{N \cdot N^{\gamma} K_{\gamma}^{1-\gamma}}{N}\right) = \Theta\left(\sqrt{N}\log^{1-\gamma} N\right)$$

and similarly for V^{odd} to I^{odd}, so by Poisson concentration we have $\Theta_{\mathbb{P}}\left(\sqrt{N}\log^{1-\gamma} N\right)$ edges between each. However for any particular $v \in V$ and $\text{par} \in \{\text{odd}, \text{even}\}$ we see a number of edges with mean

$$\frac{w(v)w(I^{\text{par}})}{w([N])} \le \frac{1 \cdot \frac{\beta}{2(1-\gamma)^2} N^{\gamma} K_{\gamma}^{1-\gamma}}{\frac{\beta}{(1-\gamma)^2} N} = \frac{\log^{1-\gamma} N}{2\sqrt{N}}$$

so the probability that this v received more than one edge from I^{par} is bounded by

$$O\left(\frac{\log^{2-2\gamma} N}{N}\right)$$

and hence by a union bound we will see only $O_{\mathbb{P}}^{\log N}(1)$ such instances. Further, any

vertex $v \in V$ has $\mathrm{d}(v) \preceq \mathrm{Pois}(1)$ and so by the union bound

$$\max_{v \in V} \mathrm{d}(v) = O_{\mathbb{P}}\left(\log N\right).$$

So because from $\Theta_{\mathbb{P}}\left(\sqrt{N}\log^{1-\gamma} N\right)$ total edges, only $O_{\mathbb{P}}^{\log N}(1)$ vertices in V^{par} have received more than 1, and at most $O_{\mathbb{P}}^{\log N}(1)$ edges at each, we conclude that $\Theta_{\mathbb{P}}\left(\sqrt{N}\log^{1-\gamma} N\right)$ vertices in V^{par} received a unique edge. Denote the sets which are connected by a unique edge $\mathcal{E} \subset V^{\mathrm{even}}$ and $\mathcal{O} \subset V^{\mathrm{odd}}$.

For the final stage of the construction, each vertex $o \in \mathcal{O}$ has conditionally

$$e(o, \mathcal{E}) \succeq \mathrm{Bin}\left(\Theta_{\mathbb{P}}\left(\sqrt{N}\log^{1-\gamma} N\right), \frac{\beta}{N}\right)$$

so we find a single edge into $\mathcal{E}$ with probability $\omega_{\mathbb{P}}\left(1/\sqrt{N}\right)$, and each vertex incident to this edge has no further edges with probability at least $1/e$. So, both have no further edges with probability at least $1/e^2$. Hence amongst the $|\mathcal{O}| = \omega_{\mathbb{P}}\left(\sqrt{N}\right)$ trials we will find an adjacent pair each of degree 2, with high probability.

We found a path $\mathcal{P}$ connecting $I^{\mathrm{odd}} \leftrightarrow V^{\mathrm{odd}} \leftrightarrow V^{\mathrm{even}} \leftrightarrow I^{\mathrm{even}}$ in the MNR model. Since each of these sets is disjoint, we know that after collapsing multi-edges to obtain the SNR model the path will still exist, and will then satisfy the criteria for our "double star". □

3.3 Diameter in the Supercritical Case

We will analyse any *rank one* Norros-Reittu graph, defined by a nonincreasing weight function

$$f : (0, 1] \to (0, \infty)$$

and vertex set $[N]/N$ in the interval. In this section, because we are interested in the existence of paths, there is no difference between the flattened (SNR) and unflattened (MNR) versions of this network.

We can assume that f is nonincreasing function without loss of generality because the purpose of the function f is to define a weight measure on $(0, \infty)$ and so can obtain a nonincreasing equivalent to any continuous function by "reordering", potentially with $f(0+) = \infty$. Implicitly in this definition we also have ellipticity

$$\min f > 0.$$

Then in the multigraph version, the number of edges between i and j is independently distributed as

$$\text{Pois}\left(\frac{1}{N} f\left(\frac{i}{N}\right) f\left(\frac{j}{N}\right)\right).$$

For example, the kernel defining our main network of interest in Definition 2.0.4 is

$$f(x) := \sqrt{\beta} x^{-\gamma}.$$

Proposition 3.3.1. *The rank one SNR model has a giant component if*

$$\int_0^1 f^2(x) \mathrm{d}x \in (1, \infty].$$

Proof. If $\int f^2 = \infty$ then we can argue exactly as in Corollary 3.0.5, taking $\epsilon \in (0,1)$ such that $\epsilon = e^{-\epsilon}$ and then containing in our rank one SNR a CL version percolated with retention probability $1 - \epsilon$. This percolated model has a giant component, which completes the argument.

If instead $1 < \int f^2 < \infty$ we apply [van der Hofstad, 2016, (6.8.13)] to say that this graph contains the GRG version with edge probabilities

$$p_{i,j} = \frac{f\left(\frac{i}{N}\right) f\left(\frac{j}{N}\right)}{N + f\left(\frac{i}{N}\right) f\left(\frac{j}{N}\right)}.$$

By [van der Hofstad, 2016, Theorem 6.10] the empirical degree distribution of this graph converges in L^1 to the mixed Poisson law $D \sim \text{Pois}(W)$ where

$$W \overset{(\mathrm{d})}{=} f\left(U_{[0,1]}\right) \int_0^1 f$$

and $U_{[0,1]}$ denotes a random variable uniformly distributed on $[0,1]$. By [van der Hofstad, 2016, Theorem 6.15] this graph is exactly uniformly distributed, conditional on its degrees, among simple graphs with these degrees.

Then [Bollobás and Riordan, 2015, Theorem 1] tells us that the uniform simple graph has a giant component when the *size-biased* limit law of its degree distribution has mean larger than 2. Recalling the Pois(λ) second moment is $\lambda + \lambda^2$, this is when

$$2 < \frac{\mathbb{E}\left(D^2\right)}{\mathbb{E}\left(D\right)} = \frac{\mathbb{E}\left(f\left(U_{[0,1]}\right)\right) + \mathbb{E}\left(f\left(U_{[0,1]}\right)^2\right) \int_0^1 f}{\mathbb{E}\left(f\left(U_{[0,1]}\right)\right)} = 1 + \int_0^1 f^2$$

which is the given condition. Then, because we have found a subgraph with a giant component, we conclude that the SNR model must also have a giant. □

Lemma 3.3.2. *Fix $p \in (0,1)$. On a sequence of graphs on $[N]$, mark each vertex with independent probability p. Then for any vertex define $D(v)$ as the minimum distance to a marked vertex. We have*

$$\max_v \left(D(v) \wedge \operatorname{diam} \mathscr{C}(v)\right) = O_{\mathbb{P}}(\log N).$$

Proof. We now construct the sets in which to observe arrivals. Any vertex v has a path from it of length at least $\operatorname{diam} \mathscr{C}(v)/2$, and so if we set

$$V_x := \{v \in [N] : \operatorname{diam} \mathscr{C}(v) \geq 2x\}$$

and give each vertex $v \in V_x$ a simple path $P(v) \ni v$ with $|P(v)| = x$

$$\mathbb{1}_{P(v)} := \mathbb{1}_{\{\text{marked vertex seen in } P(v)\}}$$

$$\mathbb{P}(D(v) < x) \geq \mathbb{E}\left(\mathbb{1}_{P(v)}\right) = 1 - (1-p)^x$$

then if we set $x = \left\lceil \frac{2 \log N}{-\log(1-p)} \right\rceil = \Theta(\log N)$ and apply the Harris inequality [Harris, 1960]

$$\begin{aligned}\mathbb{P}\left(\forall v \in V_x : \ D(v) < x\right) &\geq \mathbb{E}\left(\prod_{v \in V_x} \mathbb{1}_{P(v)}\right) \geq \prod_{v \in V_x} \mathbb{E}\left(\mathbb{1}_{P(v)}\right) \\ &= (1 - (1-p)^x)^N \geq \left(1 - \frac{1}{N^2}\right)^N \to 1.\end{aligned}$$

□

Remark 3.3.3. *A corollary of this lemma, and the idea behind the diameter proof to come, is that if we identify every marked vertex in this construction then the resultant multigraph has componentwise diameter $O_{\mathbb{P}}(\log N)$.*

To our knowledge, the following diameter result does not exist in the literature. When $\gamma < \frac{1}{2}$ we can couple to a uniform random graph and use results in [Fernholz and Ramachandran, 2007], but we require a bound for $\gamma \geq \frac{1}{2}$.

Rather than redevelop the theory, we will obtain this bound quicker by repeatedly applying the super- and sub-critical diameter theorems in [Bollobás et al., 2007].

Theorem 3.3.4. *The rank one NR network with any weight function which is supercritical*

$$\int_0^1 f^2 \in (1, \infty]$$

has componentwise diameter $O_{\mathbb{P}}(\log N)$.

Proof. The core of the argument is that we will lower bound the network model with a supercritical network that has finitely many types. Define for $M \in \mathbb{N}^+$

$$f_M(x) := f\left(\frac{\lceil Mx \rceil}{M}\right)$$

so that we have stochastic domination of the edges in the rank one networks written $\mathcal{N}f_M \preceq \mathcal{N}f$, where $\mathcal{N}g$ denotes the NR rank one network with weight function g determining the edge means before flattening.

As $M \to \infty$

$$\int_{M^{-1}}^1 f_M^2 = \frac{1}{M}\sum_{k=2}^{M} f\left(\frac{k}{M}\right)^2 \to \int_0^1 f^2 > 1$$

so that for M large enough the *truncated* model is also a supercritical rank one network, with bounded expected degree and finitely many types. In fact, if we take

$$0 < \epsilon < 1 - \frac{1}{\int_0^1 f^2}$$

then $(1-\epsilon)f_M$ is still asymptotically supercritical. Colour the mass of the measure with density f_M such that a measure with density $(1-\epsilon)f_M$ is blue and the difference measure is red:

$$f_M = (1-\epsilon)f_M + \epsilon f_M.$$

The convenience of labelling mass by primary colours is that we can colour *edges* blue if they go from blue mass to blue mass and then magenta if they go from blue mass to red mass, et cetera. The idea behind such a picture is of a Norros-Reittu network where edges of differing colours arrive independently, and so we work with multigraph NR in the proof - before flattening, they are indeed independent.

Note however that because f_M has a bounded kernel, from [van der Hofstad, 2016] we know that the simplified Norros-Reittu and Chung-Lu graphs are *asymptotically equivalent*.

Further all rank one kernels are trivially quasi-irreducible, or irreducible when restricted to the support of their weight function. We apply several theorems in [Bollobás

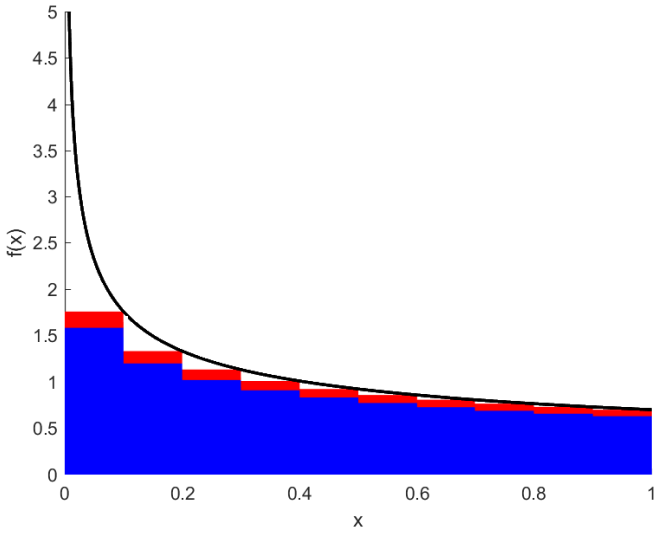


Figure 3.3: We divide the degree mass and label by red or blue, for a typical rank one scale-free network with the step count $M = 10$. Note there is some uncoloured mass.

et al., 2007] (which are stated for the equivalent Chung-Lu version) to the network of only blue edges:

- By [Bollobás et al., 2007, Theorem 3.1], because the kernel is supercritical it has a giant component.
- Because the network model has finitely many vertex types, by [Bollobás et al., 2007, Theorem 3.16], the componentwise diameter on this structure is $\Theta_{\mathbb{P}}(\log N)$.

So this subgraph of blue edges has a giant component whp, and it must be nested whp in the giant component of the edges of all colours.

After realising just the blue edges, we colour vertices in *black* if they are in the largest component of the blue edge subgraph. We now put aside the dark magenta edge mass, i.e. Poisson mass for magenta edges which would feature at least one black vertex.

Realise all the other Poisson edges in $\mathcal{N}f_M$ from mass that is not blue or dark magenta. These edges are either red or light magenta. Thinking of this as the NR multigraph, these are all independent - and questions of the diameter are the same on the multigraph or flattened version.

We finally realise the dark magenta edges, i.e. those magenta edges coming from a black vertex. Any vertex in $[N]$ and a particular black vertex are connected by a dark

magenta edge with probability at least

$$p_M := 1 - \exp\left(-\frac{1}{N}\epsilon(1-\epsilon)f(1)^2\right) = \Theta\left(\frac{1}{N}\right)$$

and so thus, after realising the number of black vertices at size $\Theta_{\mathbb{P}}(N)$, we can see vertices are incident to dark magenta edges with probability $\Theta_{\mathbb{P}}(1)$, and their incidence is entirely independent by the independence of edges for the multigraph NR network.

Applying Lemma 3.3.2, we can say that any vertex on the graph is either on a component of small diameter or within distance $O_{\mathbb{P}}(\log N)$ of a vertex of distance 1 from the blue giant. The blue giant had logarithmic diameter and so any two connected vertices in $\mathcal{N}f$ are of distance at most

$$O_{\mathbb{P}}(\log N) + 1 + O_{\mathbb{P}}(\log N) + 1 + O_{\mathbb{P}}(\log N) = O_{\mathbb{P}}(\log N)$$

by communicating through $\mathcal{N}f_M$, and we have the result. □

Lemma 3.3.5. *For the network G_N of Definition 2.0.4 with high probability, when $\beta > (1-2\gamma)^+$, every pair of two vertices in the set*

$$S := \left[\left\lfloor N \log^{-\alpha} N \right\rfloor\right],$$

where $\alpha > \frac{1}{\gamma}$, is simultaneously connected by a path of $O_{\mathbb{P}}(\log N)$ bounded weight vertices, in

$$H_\epsilon := [N] \setminus [\epsilon N],$$

for some $\epsilon \in (0,1)$ sufficiently small.

Proof. The network G_N has a Poisson number of edges between i and j, with mean

$$\beta N^{2\gamma-1} i^{-\gamma} j^{-\gamma} = \frac{1}{N}\left(\sqrt{\beta}\left(\frac{i}{N}\right)^{-\gamma}\right)\left(\sqrt{\beta}\left(\frac{j}{N}\right)^{-\gamma}\right)$$

so, in the language of this section, we can take weight function

$$f : x \mapsto \sqrt{\beta} x^{-\gamma}$$

to define precisely this model.

The subnetwork induced by H_ϵ has rank one weight function

$$f\mathbb{1}_{(\epsilon,1]}$$

and norm

$$\int_{\epsilon}^{1} f^2 > 1$$

for $\epsilon \in (0,1)$ sufficiently small.

Having fixed ϵ thus, Proposition 3.3.1 shows the induced rank one subnetwork on H_ϵ has a giant component $\mathscr{C}'$ which has $\mathscr{C}' \subseteq \mathscr{C}_{\max}$ whp. Because $|\mathscr{C}'| = \Theta_{\mathbb{P}}(N)$ we have some $\delta > 0$ such that the event

$$E_\delta = \{|\mathscr{C}'| \geq \delta N\}$$

occurs with high probability.

Now take some $s \in S$. In the MNR version of the network conditionally on E_δ we have $\text{Pois}(\mu_s)$ edges between s and $\mathscr{C}'$, where this mean has

$$\mu_s \geq \frac{1}{N} f(1) f\left(\log^{-\alpha} N\right) \cdot \delta N = \beta\delta \log^{\alpha\gamma} N.$$

Hence the probability that there is no such edge is bounded by

$$e^{-\beta\delta \log^{\alpha\gamma} N} = o\left(\frac{1}{N}\right)$$

given $\alpha\gamma > 1$. Then using $|S| \leq N$ we can connect every $s \in S$ via the union bound. The result then follows from Theorem 3.3.4.

□

Chapter 4

Interacting Particle Systems

In the previous chapter we proved results about graphs, with no reference to the occupying process. We must prove results about the processes in a general state space, which will be done in this chapter. These will combine with the previous results in later chapters.

After proving a selection of results on finite state space Markov chains which are all necessary for proofs elsewhere, we continue with a more general exposition of the voter model and contact process. The latter involves several results not absolutely required for later proofs but important for understanding the processes' behaviours.

For both models we introduce their form of duality, each being obtained by reversing time in the graphical construction (the construction from Poisson processes), and in the case of the voter model this is very useful for results on the consensus time via the coalescence time in the dual system of coalescing random walkers. When discussing the contact process, we will quickly move to a tree perspective of network exploration via local infection and then what we call the subtree contact process, which will later help to bound infection size from this local perspective.

4.1 Markov Chains

Interacting particle systems considered in this thesis are Markov chains on a finite state space, and so too are the random walkers dual to the voter model or our contact process upper bound of choice. Hence we have a few crucial results for general Markov chains which we will prove in this section.

4.1.1 Bounds on hitting times

Throughout the following two sections $X = (X_t)_{t\geq 0}$ will be a reversible, irreducible Markov chain with state space $[N] = \{1, \ldots, N\}$ and transition rates given by a generator matrix Q. Moreover, we denote by $\pi = (\pi(i))_{i\in[N]}$ the invariant measure of X.

Because the chain is irreducible, the hitting time is defined as

$$t_{\text{hit}} = \max_{k,j\in[N]} \mathbb{E}_k(T_j)$$

where $T_y := \inf\{t \geq 0 : X_t = y\}$.

For our bound on the hitting time, we will make use of the well-known correspondence between Markov chains and electric networks, see e.g. [Aldous and Fill, 2002,Levin et al., 2009]. In this context, we associate to Q a graph G_Q with vertex set $[N]$ and connect i and j by an edge, written $i \sim j$, if the conductance $c(ij)$ is nonzero, where the conductance is defined as

$$c(ij) := \pi(i)Q(i,j) = \pi(j)Q(j,i). \tag{4.1}$$

This is also known as the *ergodic flow* of the edge. Moreover, the interpretation as an electric network lets us define the effective resistance between two vertices $i, j \in [N]$, denoted $\mathcal{R}(i \leftrightarrow j)$, as in [Levin et al., 2009, Chapter 9].

To state the following proposition, we also define $\text{diam}(Q)$ to be the diameter in the graph theoretic sense for the graph obtained from Q as above. The proof uses the representation of the effective resistances in terms of the Markov chains, combined with Thomson's principle.

Proposition 4.1.1 (Conductance bounds)**.** *Let $(X_t)_{t\geq 0}$ be a reversible, irreducible Markov chain on $[N]$ with associated conductances c. Let $P_{i,j}$ be a path from i to j in G_Q and denote by $E(P_{i,j})$ the set of edges in $P_{i,j}$. Then*

$$\mathbb{E}_i(T_j) + \mathbb{E}_j(T_i) \leq \sum_{e\in E(P_{i,j})} \frac{1}{c(e)}.$$

In particular, we have

$$t_{\text{hit}} \leq \text{diam}(Q) \max_{i\sim j\in[N]} \frac{1}{c(ij)}.$$

Proof. Let $T_i^+ = \inf\{t > 0 : X_t = i, \lim_{s\uparrow t} X_s \neq i\}$ be the return time to state i.

From [Levin et al., 2009, Proposition 9.5],

$$\mathcal{R}(i \leftrightarrow j) = \frac{1}{c(i)\mathbb{P}_i(T_j < T_i^+)},$$

where $c(i) = \sum_{j\sim i} c(ij)$ is the conductance around a vertex. We further have from [Aldous and Fill, 2002, Corollary 2.8 (continuous time version)]

$$\mathbb{E}_i(T_j) + \mathbb{E}_j(T_i) = \frac{1}{\pi(i)q(i)\mathbb{P}_i(T_j < T_i^+)},$$

where $q(i) = -Q(i,i)$ is the walker speed at i, and by the choice of c these expressions are equal. Finally by Thompson's Theorem (which describes monotonicity of effective resistances with respect to edge resistances)

$$\begin{aligned}\mathbb{E}_i(T_j) + \mathbb{E}_j(T_i) = \mathcal{R}(i \leftrightarrow j) &\leq \mathcal{R}(i \leftrightarrow j \text{ through } P_{i,j}) \\ &= \sum_{\{u,v\}\in E(P_{i,j})} \frac{1}{c(uv)},\end{aligned}$$

which gives the required bound. □

4.1.2 Bounds on meeting times

In this section, we continue to use the notation from the beginning of Section 4.1.1. In particular, $(X_t)_{t\geq 0}$ is a reversible, irreducible Markov chain on $[N]$ with transition rates given by Q and invariant measure π. If $(Y_t)_{t\geq 0}$ is an independent copy of the chain with arbitrary initial condition, then the random meeting time for the two processes is

$$\tau_{\text{meet}} := \inf_{t\geq 0}\{t : X_t = Y_t\}$$

and the expected meeting time is defined by the worst case initial conditions

$$t_{\text{meet}} := \max_{x,y\in[N]} \mathbb{E}\left(\tau_{\text{meet}} \middle| X_0 = x, Y_0 = y\right).$$

It will often be easier to work with the (expected) meeting time when both chains are started in the invariant measure, i.e. we define

$$t_{\text{meet}}^{\pi} := \sum_{i,j\in[N]} \pi(i)\pi(j)\,\mathbb{E}_{i,j}(\tau_{\text{meet}}).$$

In order to make the connection to t_{meet}, we will need the time it takes to reach stationarity. There are competing definitions of the distance from stationarity, both of which are required to apply the literature results.

Definition 4.1.2. For a Markov chain on $[N]$

$$d(t) := \frac{1}{2} \max_{x \in [N]} \|p_{x,\cdot}^{(t)} - \pi(\cdot)\|_1,$$

$$\bar{d}(t) := \frac{1}{2} \max_{x,y \in [N]} \|p_{x,\cdot}^{(t)} - p_{y,\cdot}^{(t)}\|_1.$$

The mixing time t_{mix} is then defined as

$$t_{\text{mix}} := \min \left\{ t \geq 0 : d(t) \leq \frac{1}{4} \right\},$$

and the mixing time from a point $i \in [N]$ as

$$t_{\text{mix}}(i) := \min \left\{ t \geq 0 : \left\| p_{i,\cdot}^{(t)} - \pi \right\|_1 \leq \frac{1}{2} \right\}.$$

Closely related to the mixing time is the relaxation time

$$t_{\text{rel}} := \max \left\{ \frac{1}{\lambda} : \lambda \text{ a positive eigenvalue of } -Q \right\}$$

and we describe the relationship in the following lemma, as standard references are either in discrete time or using different definitions.

Lemma 4.1.3. *Write* $\pi_* := \min \pi$, *then*

$$t_{\text{rel}} \leq t_{\text{mix}} \leq 5 t_{\text{rel}} \left(1 + \frac{1}{2} \log \frac{1}{\pi_*} \right).$$

Proof. We have a similar result in [Aldous and Fill, 2002, Lemma 4.23] for the mixing time T of total variation distance $d(T) = \frac{1}{e}$

$$t_{\text{rel}} \leq T \leq t_{\text{rel}} \left(1 + \frac{1}{2} \log \frac{1}{\pi_*} \right).$$

First, because $d(t_{\text{mix}}) = \frac{1}{4} \leq \frac{1}{e}$, $t_{\text{mix}} \geq T \geq t_{\text{rel}}$ which proves the first inequality.

By the previous definitions of the mixing time and stationarity distances, we have that if $d(T) = \frac{1}{e}$, then $\bar{d}(T) \leq \frac{2}{e}$. Therefore, by the submultiplicativity of $\bar{d}$ shown

in [Aldous and Fill, 2002] for any integer $C \geq 1$ we have

$$\bar{d}(CT) \leq \left(\frac{2}{e}\right)^C$$

which is less than $\frac{1}{4}$ when $C = 5$, so $t_{\text{mix}} \leq 5T$. □

Proposition 4.1.4. *(a)*

$$t^{\pi}_{\text{meet}} \geq \frac{(1 - \sum_{i\in[N]} \pi(i)^2)^2}{4\sum_{i\in[N]} q(i)\pi(i)^2},$$

where $q(i) = -Q(i,i)$.

(b) There exists an absolute constant $c_{\text{cond}} > 0$ *such that*

$$t_{\text{meet}} \geq c_{\text{cond}} \left(\max_{A\subset[N]} \frac{\pi(A)\pi(A^c)}{\sum_{x\in A}\sum_{y\in A^c} c(xy)} \right).$$

Proof. For Part (a) see Remark 3.5 in [Chen et al., 2016].

(b) From the standard coupling bound for mixing times seen in [Aldous and Fill, 2002, Theorem 9.2] and with $\tau^{i,j}_{\text{meet}}$ as in (4.11) denoting the meeting time of walkers initially at i and j,

$$d(t) \leq \max_{i,j} \mathbb{P}\left(\tau^{i,j}_{\text{meet}} > t\right) \leq \exp\left(-\left\lfloor \frac{t}{et_{\text{meet}}} \right\rfloor\right) \leq \exp\left(1 - \frac{t}{et_{\text{meet}}}\right)$$

where the second inequality is from [Aldous and Fill, 2002, Equation (2.20)]. So by integrating

$$\frac{1}{4} t_{\text{mix}} \leq \int_0^\infty d(t)\mathrm{d}t \leq e^2 t_{\text{meet}}. \tag{4.2}$$

Because $c(xy) = \pi(x)Q(x,y)$, by [Aldous and Fill, 2002, Corollary 4.37],

$$\max_{A\subset[N]} \frac{\pi(A)\pi(A^c)}{\sum_{x\in A}\sum_{y\in A^c} c(xy)} \leq t_{\text{rel}}.$$

Combining this with Equation (4.2) and Lemma 4.1.3 proves the claim. □

In discrete time, [Peres et al., 2017, Corollary 1.2] has the consequence that, for some universal $C > 0$, $t_{\text{mix}} \leq C \min_i t_{\text{hit}}(i)$, the maximal expected hitting time of the vertex i. We present a simple proof of this fact for the convenience of the reader and to give an explicit constant factor.

Lemma 4.1.5. *For any $i \in [N]$,*

$$t_{\text{mix}}(i) \leq 2\mathbb{E}_\pi\left(T_i\right).$$

Proof. Let $i \in [N]$, then by Cauchy-Schwarz we have that

$$\left\|p_{i\cdot}^{(t)} - \pi\right\|_1^2 = \sum_{j\in[N]} \left|\frac{p_{ij}^{(t)}}{\pi(j)} - 1\right|\pi(j) \leq \left(\sum_{i\in[N]} \left|\frac{p_{ij}^{(t)}}{\pi(j)} - 1\right|^2 \pi(j)\right)^2 = \left\|\frac{p_{i\cdot}^{(t)}}{\pi} - 1\right\|_\pi^2.$$

To simplify the right hand side, we use reversibility to obtain

$$\left\|\frac{p_{i\cdot}^{(t)}}{\pi} - 1\right\|_\pi^2 = -1 + \sum_j \frac{\left(p_{ij}^{(t)}\right)^2}{\pi(j)} = -1 + \frac{1}{\pi(i)}\sum_j p_{ij}^{(t)} p_{ji}^{(t)} = -1 + \frac{p_{ii}^{(2t)}}{\pi(i)}.$$

Now, by [Aldous and Fill, 2002, Lemma 2.11], we have that for any $t \geq 0$,

$$\mathbb{E}_\pi\left(T_i\right) = \int_0^\infty \left(-1 + \frac{p_{ii}^{(s)}}{\pi(i)}\right) \mathrm{d}s \geq 2t\left(-1 + \frac{p_{ii}^{(2t)}}{\pi(i)}\right),$$

because the integrand is non-increasing [Aldous and Fill, 2002, Equation 3.40]. Combining these inequalities, we have for $t > 0$,

$$\left\|p_{i\cdot}^{(t)} - \pi\right\|_1 \leq \left\|\frac{p_{i\cdot}^{(t)}}{\pi} - 1\right\|_\pi \leq \sqrt{\frac{\mathbb{E}_\pi\left(T_i\right)}{2t}}. \tag{4.3}$$

Hence, if t is such that

$$\frac{\mathbb{E}_\pi\left(T_i\right)}{2t} \leq \frac{1}{4},$$

then we can deduce that $t_{\text{mix}}(i) \leq t$, which completes the proof. □

Proposition 4.1.6. *For two independent copies $(X_t)_t$ and $(Y_t)_t$ of a Markov Chain on $[N]$, and any state $s \in [N]$, we find*

$$t_{\text{mix}} \leq 16\, t_{\text{hit}}(s)$$

and further we can construct a time for the product chain with

$$\mathbb{E}_{x,y}(\mathcal{S}) \leq 188\, t_{\text{hit}}(s)$$

which is a strong stationary time in the sense that for any $t \geq 0$ we have $\mathcal{L}(X_{t+\mathcal{S}}, Y_{t+\mathcal{S}}) =$

$\pi \otimes \pi$ *and, further,* $(X_{t+S}, Y_{t+S})_{t\geq 0}$ *and* $\mathcal{S}$ *are independent.*

Proof. Define the time $M_1 = 8t_{\text{hit}}(s)$. Then, by Markov's inequality

$$\max_x \mathbb{P}_x(T_s \geq M_1) \leq \max_x \frac{\mathbb{E}_x(T_s)}{M_1} = \frac{t_{\text{hit}}(s)}{M_1} = \frac{1}{8}$$

so we will hit s in the timeframe $[0, M_1]$ with probability at least $\frac{7}{8}$. Define also the time M_2 by

$$\frac{\mathbb{E}_\pi(T_s)}{2M_2} = \frac{1}{16}$$

then by recalling equation (4.3) we have that

$$\frac{1}{2}\|p_{s,\cdot}^{(M_2)} - \pi(\cdot)\|_1 \leq \frac{1}{8}.$$

Hence, by distinguishing the cases of hitting s by M_1, or not, we obtain that $d(M_1 + M_2) \leq \frac{1}{4}$. Thus,

$$t_{\text{mix}} \leq M_1 + M_2 = 8\, t_{\text{hit}}(s) + 8\, \mathbb{E}_\pi(T_s) \leq 16\, t_{\text{hit}}(s).$$

It remains to prove the second claim. By Theorem 1.1 in [Fill, 1991] we can construct a strong stationary time with

$$\mathbb{P}_s(\mathcal{S}_X > t) = \text{sep}_s(t) =: 1 - \min_{j\in[N]} \frac{p_{s,j}^{(t)}}{\pi(j)}.$$

Then we recover several definitions and results from [Aldous and Fill, 2002] given in Definition 4.1.2. These various definitions of distance from stationarity satisfy

$$\text{sep}_s(2t) \leq \max_{v\in[N]} \text{sep}_v(2t) < 2\,\bar{d}(t) \leq 4\, d(t).$$

Therefore, we have that

$$\tau_1 := \min\left\{t : \bar{d}(t) \leq \frac{1}{2}\right\} \leq \min\left\{t : d(t) \leq \frac{1}{4}\right\} = t_{\text{mix}}.$$

Then we use that $\bar{d}$ is submultiplicative to obtain

$$\bar{d}(t) \leq 2^{-\lfloor t/\tau_1 \rfloor} \leq 2^{-\lfloor t/t_{\text{mix}} \rfloor}.$$

Thus, we can bound the expectation of the time to stationarity

$$\mathbb{E}_s(\mathcal{S}_X) = 2\int_0^\infty \mathrm{sep}_s(2t)\mathrm{d}t \leq 4\int_0^\infty 2^{1-t/t_{\mathrm{mix}}}\mathrm{d}t = \frac{8\,t_{\mathrm{mix}}}{\log 2} \leq \frac{64\,t_{\mathrm{hit}}(s)}{\log 2}.$$

This becomes a strong stationary time for $(X_t)_t$ with $X_0 = x$ by constructing another time $\tilde{\mathcal{S}}_X$ which simply waits for the event when the walker hits s, and then waits for $\mathcal{S}_X$. Thus

$$\mathbb{E}_x\left(\tilde{\mathcal{S}}_X\right) \leq t_{\mathrm{hit}}(s) + \frac{64\,t_{\mathrm{hit}}(s)}{\log 2} < 94\,t_{\mathrm{hit}}(s).$$

We construct the symmetric time $\tilde{\mathcal{S}}_Y$ for $(Y_t)_t$ and then finally our object is the time

$$\mathcal{S} := \tilde{\mathcal{S}}_X \vee \tilde{\mathcal{S}}_Y$$

so that

$$\mathbb{E}_{x,y}(\mathcal{S}) \leq \mathbb{E}_x\left(\tilde{\mathcal{S}}_X\right) + \mathbb{E}_y\left(\tilde{\mathcal{S}}_Y\right) \leq 188\,t_{\mathrm{hit}}(s),$$

as claimed. □

Proposition 4.1.7. *For any state $s \in [N]$*

$$t_{\mathrm{meet}} \leq \frac{189\,t_{\mathrm{hit}}(s)}{\pi(s)}.$$

Proof. From any configuration of two walkers, we can apply Proposition 4.1.6 to construct a strong stationary time $\mathcal{S}$ with $\mathbb{E}(\mathcal{S}) \leq 188\,t_{\mathrm{hit}}(s)$. Then, wait for $(X_t)_t$ to hit s, which in expectation takes an additional time period of length $t_{\mathrm{hit}}(s)$.

On $(X_t)_t$ hitting s, $(Y_t)_t$ is still in independent stationarity, and so we have exactly probability $\pi(s)$ to meet at that instant. Otherwise, we restart the argument with mixing and hitting periods to get another chance to meet at s.

Thus we have to repeat the attempt no more than $\mathrm{Geom}(1/\pi(s))$ times, and each attempt conditionally expects to take no longer than $188\,t_{\mathrm{hit}}(s) + t_{\mathrm{hit}}(s)$. □

In this proof we used this slightly over-powerful machinery of the constructed strong stationary time to avoid having to introduce a mixing time of total variation distance less than $\frac{1}{4}$. We comment, however, that the independence given by this time was not really necessary for the proof.

Remark 4.1.8. *We find the following illustrative discrete-time bound in [Kanade et al., 2016]*

$$t^{\pi}_{\text{meet}} = O\left(\frac{t_{\text{mix}}}{||\pi||_2^2}\right)$$

which, while appearing a better bound, is commonly not so for Markov chains on trees. The mixing time for a Markov chain on a tree (which must always be a reversible chain) is always the hitting time of a central *vertex, i.e. one with*

$$\mathbb{E}_\pi T_c = \min_{v\in[N]} \mathbb{E}_\pi T_v.$$

Then,

$$t_{\text{mix}} = \Theta\left(t_{\text{hit}}(c)\right)$$

and so because $\|\pi\|_2^2 \leq \|\pi\|_\infty$, Proposition 4.1.7 will often give a tighter bound.

In the following, we will need the following large deviations result given in [Saloff-Coste, 1997].

Theorem 4.1.9. *For any finite, irreducible continuous-time Markov chain $(X_t)_t$ with initial stationary distribution π, and any function on the state space f with*

$$\langle f, \pi\rangle = 0, \qquad ||f||_\infty \leq 1,$$

we have for $x > 0$

$$\mathbb{P}_\mu\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s > x\right) \leq ||\mu/\pi||_2 \exp\left(-\frac{x^2 t}{10\, t_{\text{rel}}}\right),$$

where μ is an arbitrary distribution on the state space.

We now use the concept of the chain $(X_t)_{t\geq 0}$ *observed on a subset* $V \subset N$ described in Section 2.7.1 in [Aldous and Fill, 2002]: define a clock process

$$U(t) := \int_0^t \mathbb{1}_V\left(X_s\right)\mathrm{d}s,$$

with generalised right-continuous inverse U^{-1}. Then the partially observed chain $(P_t)_{t\geq 0}$ is defined for any $t \geq 0$ via

$$P_t := X_{U^{-1}(t)}.$$

This corresponds to the deletion of states in V^c from the history of $(X_t)_{t\geq 0}$, and so

it can be shown that $(P_t)_{t\geq 0}$ is Markovian and has the natural stationary distribution

$$\frac{\pi(\cdot)\mathbb{1}_V(\cdot)}{\pi(V)}.$$

Then, we can define the random *subset meeting time* $\tau^{\pi}_{\text{meet}}(A)$ analogously to τ^{π}_{meet} except for the partially observed product chain on $A \times A$ rather than the full chain. Similarly, $t^{\pi}_{\text{meet}}(A) = \mathbb{E}(\tau^{\pi}_{\text{meet}}(A))$.

Theorem 4.1.10. *For any $A \subset [N]$,*

$$t_{\text{meet}} \leq 188 t_{\text{hit}}(s) + \frac{2\, t^{\pi}_{\text{meet}}(A)}{\pi(A)^2} + \frac{640\, t_{\text{hit}}(s)}{\pi(A)^4}.$$

Proof. We first prove the claim

$$t^{\pi}_{\text{meet}} \leq \frac{2\, t^{\pi}_{\text{meet}}(A)}{\pi(A)^2} + \frac{40\, t_{\text{mix}}}{\pi(A)^4}. \tag{4.4}$$

Consider two independent copies $(X_t, Y_t)_t$ of the stationary chain, such that in particular, for any $t \geq 0$ we have that $\mathcal{L}(X_t, Y_t) = \pi \otimes \pi$. Define the time-change

$$U(t) := \int_0^t \mathbb{1}_{A\times A}\left(X_s, Y_s\right) \mathrm{d}s.$$

Then, the product chain $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ observed on $A\times A$ satisfies $(\tilde{X}_t, \tilde{Y}_t) = (X_{U^{-1}(t)}, Y_{U^{-1}(t)})$ for any $t \geq 0$. Therefore, we have that for any $t \geq 0$,

$$\mathbb{P}(U(\tau^{\pi}_{\text{meet}}) \geq t) \leq \mathbb{P}(\tau^{\pi}_{\text{meet}}(A) \geq t),$$

since a meeting might also happen outside A. In particular, we can deduce that

$$\begin{aligned}
\mathbb{E}(\tau^{\pi}_{\text{meet}}) &= \int_0^{\infty} \mathbb{P}(\tau^{\pi}_{\text{meet}} > t)\, \mathrm{d}t \\
&\leq \int_0^{\infty} \mathbb{P}\left(U(\tau^{\pi}_{\text{meet}}) > U(t); U(t) \geq \frac{\pi(A)^2}{2} t\right) \mathrm{d}t + \int_0^{\infty} \mathbb{P}\left(U(t) \leq \frac{\pi(A)^2}{2} t\right) \mathrm{d}t \\
&\leq \int_0^{\infty} \mathbb{P}\left(\tau^{\pi}_{\text{meet}}(A) \geq \frac{\pi(A)^2}{2} t\right) \mathrm{d}t + \int_0^{\infty} \mathbb{P}\left(U(t) \leq \frac{\pi(A)^2}{2} t\right) \mathrm{d}t \\
&\leq \frac{2}{\pi(A)^2} t^{\pi}_{\text{meet}}(A) + \int_0^{\infty} \mathbb{P}\left(U(t) \leq \frac{\pi(A)^2}{2} t\right) \mathrm{d}t.
\end{aligned} \tag{4.5}$$

It remains to estimate the second integral on the right hand side.

For this purpose, we apply Theorem 4.1.9 with $f := \mathbb{1}_{A\times A} - \pi(A)^2$ to obtain

$$\mathbb{P}_\pi\left(\frac{1}{t}\int_0^t \mathbb{1}_{A\times A}(X_s, Y_s)\,\mathrm{d}s - \pi(A)^2 < -x\right) \le \exp\left(-\frac{x^2 t}{10\, t_{\mathrm{rel}}}\right)$$

and hence

$$\mathbb{P}\left(U(t) \le \frac{\pi(A)^2}{2}t\right) = \mathbb{P}\left(\frac{U(t)}{t} - \pi(A)^2 \le -\frac{\pi(A)^2}{2}\right) \le \exp\left(-\frac{t\pi(A)^4}{40\, t_{\mathrm{rel}}}\right).$$

We deduce that

$$\int_0^t \mathbb{P}\left(U(t) \le \frac{\pi(A)^2}{2}t\right) \le \frac{40\, t_{\mathrm{rel}}}{\pi(A)^4}.$$

Moreover, Lemma 4.1.3 gives $t_{\mathrm{rel}} \le t_{\mathrm{mix}}$ which combined with (4.5) gives the claim (4.4).

To obtain the statement of the theorem, recall from Proposition 4.1.6 that there exists a strong stationary time $\mathcal{S}$ such that

$$t_{\mathrm{mix}} \le 16\, t_{\mathrm{hit}}(s) \quad \text{and} \quad \mathbb{E}_{x,y}(\mathcal{S}) \le 188\, t_{\mathrm{hit}}(s).$$

Using the stationary time constructed in this corollary gives the bound

$$t_{\mathrm{meet}} \le \max_{x,y\in[N]} \mathbb{E}_{x,y}(\mathcal{S}) + t^{\pi}_{\mathrm{meet}},$$

which together with (4.4) proves the theorem. □

4.1.3 Metastability

The metastable version of a chain on V with generator matrix $Q = [q_{ij}]$, called quasistationary in [Aldous and Fill, 2002, Section 3.6.5], is obtained by conditioning the dynamics so that the chain does not hit some set $A \subset V$. If this conditional dynamic has a unique stationary distribution, it is the metastable distribution α and satisfies the equations

$$\begin{aligned} \alpha(j) &= 0 \quad \text{if } j \in A \\ \sum_{i\in A^c} \alpha(i) q_{ij} &= -\rho\alpha(j) \quad \text{if } j \in A^c \\ \sum_{j\in A^c} \alpha(j) &= 1 \end{aligned} \tag{4.6}$$

for some exit rate parameter $\rho > 0$. If we write Q^A for the generator Q restricted to A^c then [Aldous and Fill, 2002] tells us that these equations are always uniquely solvable if the generator Q^A defines an irreducible Markov chain, and in fact ρ is the spectral radius of Q^A.

In balancing the exit rates at every state, we have found a distribution α from which, with the unconditional Markov chain defined by Q, the hitting time of the set A is precisely exponentially distributed with rate parameter ρ (see [Aldous and Fill, 2002, Section 3.6.5]).

We have one relatively intuitive proposition about ergodic flow travelling towards higher escape rates, which will help in later applications of this theory.

Proposition 4.1.11. *Consider a continuous time Markov chain on state space $V = S \cup \{\dagger\}$ with rate matrix Q such that $Q^\dagger$ is irreducible on S. For each $s \in S$, write $r(s) := Q(s, \dagger)$ for the escape rate. Then if S can be partitioned into $S = A \cup B$ with*

$$\sup_{a \in A} r(a) \leq \inf_{b \in B} r(b)$$

then we find the metastable distribution α, obtained by conditioning to not hit $\dagger$, satisfies

$$\sum_{a \in A} \alpha(a) Q(a, B) \geq \sum_{b \in B} \alpha(b) Q(b, A).$$

Proof. α denotes the solution to the metastable problem as stated in the proposition. We can interpret α, using the equations (4.6), as stationary with respect to the dynamic which places walkers at a random location distributed according to α after their escape to $\dagger$. To be clear, this is the rate matrix $\tilde{Q}$ on S with

$$\tilde{Q}(i, j) = Q(i, j) + r(i)\alpha(j)$$

for any $i, j \in S$. Then we can *collapse* this Markov chain as in [Aldous and Fill, 2002, Section 2.7.3] to the two state chain on $\{A, B\}$ with exponential rates

$$\tilde{q}_{AB} = \sum_{a \in A} \frac{\alpha(a)}{\alpha(A)} Q(a, B) + \sum_{a \in A} \frac{\alpha(a)}{\alpha(A)} r(a)\alpha(B)$$
$$\tilde{q}_{BA} = \sum_{b \in B} \frac{\alpha(b)}{\alpha(B)} Q(b, A) + \sum_{b \in B} \frac{\alpha(b)}{\alpha(B)} r(b)\alpha(A)$$

which has stationary distribution $\alpha(A)$, $\alpha(B)$ because it is the collapsed version of a

chain with stationary distribution α. Checking detailed balance for this two-state chain,

$$\sum_{a\in A}\alpha(a)\left(Q(a,B)+r(a)\alpha(B)\right)=\sum_{b\in B}\alpha(b)\left(Q(b,A)+r(b)\alpha(A)\right)$$

and hence

$$\sum_{a\in A}\alpha(a)Q(a,B)+\alpha(A)\alpha(B)\sup_{a\in A}r(a)\geq\sum_{b\in B}\alpha(b)Q(b,A)+\alpha(A)\alpha(B)\inf_{b\in B}r(b).$$

Recalling from the claim that $\sup_{a\in A}r(a)\leq\inf_{b\in B}r(b)$, we conclude

$$\sum_{a\in A}\alpha(a)Q(a,B)\geq\sum_{b\in B}\alpha(b)Q(b,A).$$

□

4.2 Voter Models

It was used in the original paper [Clifford and Sudbury, 1973], and has been well-known since [Liggett, 1985], that voter models are dual to coalescing random walks which trace back where opinions came from. In particular on an arbitrary finite state space, if we consider the final coalescence time at which a set of random walks started at each vertex in the underlying graph have coalesced into a single walker, then this upper bounds the consensus time and thus any result that bounds the expectation of this coalescence time [Oliveira, 2013, Kanade et al., 2016] also bounds the expected consensus time in the voter model. In this section we will describe this duality in more detail, which is the main tool for analysis of the voter model, and also give the useful consequences.

The duality can be described via a graphical construction. We start with the graph $\{(j,t)\,:\,j\in[N],t\geq 0\}$ and independent Poisson point processes $(V_{i,j}(t))_{t\in\mathbb{R}},i\neq j$ (with rates $Q(i,j)$ respectively). If t_k denotes a jump of $V_{i,j}$ we draw an arrow from (t_k,j) to (t_k,i). Given any initial condition $\eta_0\in O^{[N]}$, we then let the opinions flow upwards starting at time 0 and if they encounter an arrow following the direction of the arrow and replacing the opinion of the voter found at the tip of the arrow. Now, we fix a time horizon $T>0$ and we start with N random walkers located at each of the points $(j,T),j\in[N]$, then the trajectories follow the graph downwards, following each arrow and if two walkers meet they coalesce. Denote by $\xi_t^T\subseteq[N]=\{\xi_t^T(j)\,,\,j\in[N]\}$ the set of positions of these walkers at time $t\geq 0$, where thus $\xi_0^T=[N]$. From this construction,

it follows that each walker follows the dynamics of the Markov chain X, so we obtain a system of coalescing Markov chains/random walks. Moreover, one can immediately see that the voter model at time T can be obtained, by tracing the random walk paths backwards, i.e. for any $j \in [N]$,

$$\eta_T(j) = \eta_0(\xi_T^T(j)). \tag{4.7}$$

The "tracing" in this construction gives an explicit coupling at every point in time which means that we are actually describing a *strong pathwise dual* in the sense of [Jansen et al., 2014].

We are interested in general reversible Markov chains, so we do not necessarily assume that the Markov chain is irreducible. However, since $X = (X_t)_{t\geq 0}$ is reversible, we can decompose the state space into its irreducible components, which we will denote by $C_1, \ldots, C_k$, so that X restricted to C_j is irreducible. In this case, for any $j \in [k]$, we denote the consensus time on the jth component by

$$\tau_{\text{cons}}(C_j) := \inf\{t \geq 0 \,:\, \eta_t|_{C_j} \text{ is constant}\}.$$

Then, as in Definition 2.1.2 we define the consensus time

$$\tau_{\text{cons}} := \max_{j\in[k]} \tau_{\text{cons}}(C_j).$$

Our main interest in this thesis is in the case when $O = \{0, 1\}$ and the initial conditions η_0 are distributed according to μ_u, the product of Bernoulli(u) measures for some $u \in [0, 1]$. Then, we set

$$t_{\text{cons}}^{(u)} := \mathbb{E}_{\mu_u}(\tau_{\text{cons}}).$$

For the duality, it will be easier to consider the voter model where each voter starts with a different opinion, i.e. $\eta_0 = [N]$. Here, we define

$$t_{\text{cons}}^{*} := \mathbb{E}_{[N]}(\tau_{\text{cons}}).$$

For the system of coalescing random walks, we define for each irreducible component $C_j, j \in [k]$,

$$\tau_{\text{coal}}(C_j) := \inf\{t \geq 0 \,:\, |\xi_t^T| = 1\},$$

i.e. the first time all walkers in this component have coalesced into a single walker. This coalescence may occur after the time horizon T, but that does not cause issues with our

construction as the Poisson processes $V_{i,j}$ were generated on $\mathbb{R}$, and so we can continue the duality into jumps that were in negative time for the original voter model. Moreover, we then define

$$t_{\text{coal}} := \mathbb{E}_{[N]}(\tau_{\text{coal}}), \quad \text{where } \tau_{\text{coal}} := \sup_{j\in[k]} \tau_{\text{coal}}(C_j).$$

By duality, we have that if the voter model starts in $\eta_0 = [N]$, then

$$\mathbb{P}_{[N]}(\tau_{\text{coal}} \leq T) = \mathbb{P}_{[N]}(\tau_{\text{cons}} \leq T),$$

so τ_{coal} and τ_{cons} agree in distribution and in particular $t_{\text{coal}} = t^*_{\text{cons}}$. For opinion sets with fewer than N distinct opinions, we can map the distinct opinions model to the reduced set to see that this consensus time upper bounds one with $\eta_0 \in \{0,1\}^{[N]}$.

The first basic property of the voter model is that the opinion proportions have a martingale property, as proven in e.g. [Cooper and Rivera, 2016, Lemma 6].

Proposition 4.2.1. *Suppose a voter model $(\eta_t)_t$ in $\{0,1\}^N$ on the graph $G = ([N], E)$ has a dual Markov chain with an invariant distribution π. Then*

$$M_t := \sum_{v=1}^{N} \eta_t(v)\pi(t), \quad t \geq 0,$$

defines a martingale in $[0,1]$.

In [Chen et al., 2016] they discuss convergence of a wide class of well-mixed voter models, which all have convergence of this martingale to the Wright-Fisher diffusion [Etheridge, 2011] as the graph size diverges $N \to \infty$.

Definition 4.2.2 (Wright-Fisher diffusion)**.** The Wright-Fisher diffusion $(Y, \mathbb{P}_u)$ is the continuous martingale in $[0,1]$ with

$$\left.\begin{aligned} Y_0 &= u \\ \mathrm{d}\langle Y\rangle_t &= Y_t(1-Y_t)\mathrm{d}t \end{aligned}\right\} \quad \text{a.s.}$$

This means that the limiting distribution shape distribution of these well-mixed voter models is universal. The following proposition identifies the "coalescent" shape of this distribution, and explains why this shape distribution is natural for consensus from the independent Bernoulli initial condition μ_u.

Proposition 4.2.3. *Construct a Wright-Fisher diffusion $(Y_t)_t$ with $Y_0 = u$. If $(Z_m)_m$ is an independent sequence of* Exp(1) *random variables, and G is independently distributed*

on $\{k \geq 1\}$ *with*

$$\mathbb{P}(G = k) = u^k(1-u) + (1-u)^k u$$

then we identify an equality in distribution of τ^Y_{exit}*, the first hitting time of* 0 *or* 1 *for the Wright-Fisher diffusion,*

$$\tau^Y_{\text{exit}} \overset{\text{(d)}}{=} \sum_{m>G} \frac{2}{m(m-1)} Z_m.$$

Proof. Consider the standard voter model η on the complete graph on $[N]$. From [Chen et al., 2016, Proposition 2.6] it follows that $\forall u \in (0,1)$

$$\left(\frac{\tau^{(N)}_{\text{cons}}}{t^{\pi}_{\text{meet}}}, \mathbb{P}^{(n)}_{\mu_u} \right) \overset{\text{(d)}}{\to} \left(\tau^Y_{\text{exit}}, \mathbb{P}_u \right)$$

where here μ_u is the independent Bernoulli initial condition on the complete graph.

From [Etheridge, 2011, Example 3.21] we have

$$\mathbb{E}_u \left(\tau^Y_{\text{exit}} \right) = 2u \log \frac{1}{u} + 2(1-u) \log \frac{1}{1-u}, \tag{4.8}$$

and by identifying the geometric distribution for the meeting time of two stationary walkers we claim

$$t^{\pi}_{\text{meet}} = \frac{N-1}{2}.$$

Now consider the natural coupling of the voter model η with the pathwise dual, the coalescent CSRW walkers ζ which step $i \to j$ whenever i adopts the opinion of j in η. It follows from the duality described in this section that

$$\mathbb{P}\left(\tau^{(N)}_{\text{cons}} < t \middle| \eta_0 = \mathbb{1}_A \right) = \mathbb{P}(\zeta^{[N]}_t \subseteq A) + \mathbb{P}(\zeta^{[N]}_t \subseteq A^c) \tag{4.9}$$

and further, we can couple $(\zeta^{[N]}_t)_t$ pathwise with $(\zeta^A_t)_t$ for $A \subseteq [N]$ such that they are both still coupled to η. Due to the symmetry of the complete graph we can replace terminal positions with initial positions

$$\mathbb{P}(\zeta^{[N]}_t \subseteq A) = \mathbb{P}(\zeta^{[N]}_t = \zeta^A_t). \tag{4.10}$$

By combining Equations 4.9 and 4.10 we see that

$$\mathbb{P}(\tau^{(N)}_{\text{cons}} < t) = \mathbb{P}(\{\zeta^{[N]}_t = \zeta^A_t\} \text{ or } \{\zeta^{[N]}_t = \zeta^{A^c}_t\})$$

and so in fact the distribution of the voter model consensus time is precisely the distri-

bution of the time at which all remaining particles in the coalescent process ζ had the same initial opinion in η. Also, opinions of their initial positions in η are completely independent according to μ_u. Thus the number of particles still remaining when we hit $\tau_{\text{cons}}^{(N)}$ is independent of ζ, with distribution $G \wedge N$.

We observe finally that any two particles on the complete graph will coalesce at rate $2/N$, and so $m \leq N$ particles will coalesce at rate $\frac{2}{N}\binom{m}{2}$. Writing $(Z_m)_{m=2}^{\infty}$ for independent exponential random variables with rate 1:

$$\frac{\tau_{\text{cons}}^{(N)}}{t_{\text{meet}}^{\pi}} \overset{\text{(d)}}{=} \frac{N}{N-1} \sum_{m=G\wedge N+1}^{N} \frac{2}{m(m-1)} Z_m \overset{\text{(L}^2\text{)}}{\to} \sum_{m=G+1}^{\infty} \frac{2}{m(m-1)} Z_m$$

as $n \to \infty$. Thus we have identified

$$\tau_{\text{exit}}^{Y} \overset{\text{(d)}}{=} \sum_{m>G} \frac{2}{m(m-1)} Z_m.$$

□

We can now use that this implies the exit time is dominated by the coalescent with $G = 1$, taking $t < 1$, to find

$$\mathbb{E}\left(e^{t\tau_{\text{exit}}^{Y}}\right) \leq \prod_{m=2}^{\infty} \left(\frac{1}{1-\frac{2t}{m(m-1)}}\right) = -2\pi t \sec\left(\frac{1}{2}\pi\sqrt{8t+1}\right)$$

which is finite in an interval around 0. Apply Markov's Inequality to see

$$\mathbb{P}\left(\tau_{\text{exit}}^{Y} > x\right) \leq 7 \cdot 2^{-x}$$

so we have *subexponential tails* for this consensus time shape, after rescaling by the mean $t_{\text{cons}}^{(u)}$. As the following lemma shows, we can also get bounds for this scale $t_{\text{cons}}^{(u)}$ which is the main object of interest in this thesis.

Lemma 4.2.4 (Duality)**.** *In the setting above we have, for any $u \in (0,1)$,*

$$t_{\text{cons}}^{(u)} \leq t_{\text{coal}}.$$

Suppose additionally that the dual Markov chain is irreducible, then for all $u \in (0,1)$,

$$t_{\text{cons}}^{(u)} \geq 2u(1-u)\, t_{\text{coal}}.$$

Proof. By recolouring we can see that we reach consensus from the product Bernoulli

measure μ_u before we do from unique colours, and hence

$$t_{\text{cons}}^{(u)} \leq t_{\text{cons}}^{*} = t_{\text{coal}}.$$

For the other direction, suppose that the dual Markov chain is irreducible. Then, observe from the duality relation (4.7) that

$$\mathbb{P}_{\mu_u}\left(\eta_T \text{ constant}\right) = \mathbb{P}\left(\mu_u \text{ constant on } \xi_T^T\right) = \mathbb{E}\left(u^{|\xi_T^T|} + (1-u)^{|\xi_T^T|}\right)$$

which we can crudely upper bound by considering the event $\{|\xi_T^T| = 1\}$

$$\begin{aligned}\mathbb{E}\left(u^{|\xi_T^T|} + (1-u)^{|\xi_T^T|}\right) &\leq \mathbb{P}\left(|\xi_T^T| = 1\right) + \left(u^2 + (1-u)^2\right)\mathbb{P}\left(|\xi_T^T| \geq 2\right) \\ &= 1 - 2u(1-u)\mathbb{P}\left(|\xi_T^T| \geq 2\right).\end{aligned}$$

Therefore we have

$$\begin{aligned}t_{\text{cons}}^{(u)} &= \int_0^\infty \mathbb{P}_{\mu_u}(\tau_{\text{cons}} \geq t)\,\mathrm{d}t \\ &= \int_0^\infty 1 - \mathbb{E}\left(u^{|\xi_T^T|} + (1-u)^{|\xi_T^T|}\right)\mathrm{d}T \\ &\geq 2u(1-u)\int_0^\infty \mathbb{P}\left(|\xi_T^T| \geq 2\right)\mathrm{d}T = 2u(1-u)t_{\text{coal}},\end{aligned}$$

where we used irreducibility in the last step. □

We will control the time t_{coal} until all random walkers have coalesced using the following two bounds in terms of two auxiliary quantities that we define next. First of all, let $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ be two independent reversible Markov chains with generator Q. Then, define the meeting time for the component numbered $j \in [k]$ as

$$t_{\text{meet}}(C_j) = \max_{x,y\in C(j)} \mathbb{E}_{x,y}(\tau_{\text{meet}}), \quad \text{where } \tau_{\text{meet}} = \inf\{t \geq 0 \,:\, X_t = Y_t\}.$$

Moreover, an important role will be played by the hitting time defined for the component numbered $j \in [k]$ as

$$t_{\text{hit}}(C_j) = \max_{x,y\in C_j} \mathbb{E}_x(T_y), \quad \text{where } T_y = \inf\{t \geq 0 \,:\, X_t = y\}.$$

Both quantities give bounds on the coalescence time and thus on the consensus time.

Proposition 4.2.5. *With the notation as above, we have that*

$$\sup_{j\in[k]} t_{\text{meet}}(C_j) \leq t_{\text{coal}} \leq e(2+\log N) \sup_{j\in[k]} t_{\text{meet}}(C_j).$$

Moreover, for any $j \in [k]$,

$$t_{\text{meet}}(C_j) \leq t_{\text{hit}}(C_j).$$

Remark 4.2.6. *Recall that t_{coal} is defined as*

$$t_{\text{coal}} = \mathbb{E}_{[N]}\left(\sup_{j\in[k]} \tau_{\text{coal}}(C_j)\right)$$

so the non-standard part of the statement is that we can take the supremum out of the expectation. For irreducible chains, the statement is Proposition 14.11 in [Aldous and Fill, 2002]. However, their proof does not really need this extra assumption. For the convenience of the reader, we repeat the proof below. Furthermore, we note that for reducible chains the first bound is shown in [Oliveira, 2012] without the $\log N$ *factor. The stronger bound does not hold without the assumption of irreducibility. Indeed, by looking at a Markov chain with N components of size 2 each (e.g. with transition rates 1 within these components), it becomes obvious that the factor* $\log N$ *in the proposition is sharp.*

Proof. The reversible Markov chain decomposes into irreducible recurrence classes - write $\mathscr{C}(i)$ for the class containing the state i. As in the proof of [Aldous and Fill, 2002, Proposition 14.11], consider a walker $W^{(i)}$ independently started in i. We have $N(N+1)/2$ meeting times

$$\tau^{i,j}_{\text{meet}} := \inf\left\{t \geq 0 : W^{(i)}_t = W^{(j)}_t\right\} \tag{4.11}$$

for the walkers $1 \leq i \leq j \leq N$, where we define $\inf \emptyset := \infty$ and $\tau^{i,i} := 0$. Define a function f which maps all elements in a recurrence class $\mathscr{C}(i)$ to a label $\min \mathscr{C}(i)$ which is of lowest index in that component.

$$\text{f} : i \mapsto \min \mathscr{C}(i)$$

Then we can construct the coalescing walker from independent walkers by killing the walker of larger initial position at any meeting event, which we think of as making it follow the vertex of smaller initial position. Thus we can say, for the non-independent

walker meeting times obtained in this construction,

$$\tau_{\text{coal}} := \max_{i=1}^{N} \tau_{\text{coal}}(\mathscr{C}(i)) \leq \max_{i=1}^{N} \tau_{\text{meet}}^{i,f(i)}.$$

We then apply a result for the general exponential tails of hitting times of finite Markov chains [Aldous and Fill, 2002, Equation 2.20]: from arbitrary initial distribution μ and for a continuous time reversible chain, for any subset $A \subset V$

$$\mathbb{P}_\mu(T_A > t) \leq \exp\left(-\left\lfloor \frac{t}{e \max_v \mathbb{E}_v T_A} \right\rfloor\right).$$

For the meeting time variables, which are hitting times for the product chain, this leads to

$$\mathbb{P}(\tau_{\text{meet}}^{i,j} > t) \leq \exp\left(-\left\lfloor \frac{t}{e t_{\text{meet}}} \right\rfloor\right).$$

We can deduce by the union bound that

$$\mathbb{P}(\tau_{\text{coal}} > t) \leq \sum_{i=1}^{N} \mathbb{P}\left(\tau_{\text{meet}}^{f(i),i} > t\right) \leq N \exp\left(-\left\lfloor \frac{t}{e t_{\text{meet}}} \right\rfloor\right).$$

Finally, we integrate as in [Aldous and Fill, 2002, Proposition 14.11] to get

$$t_{\text{coal}} \leq \int_0^\infty 1 \wedge \left(Ne \exp\left(-\frac{t}{e t_{\text{meet}}}\right)\right) \mathrm{d}t = e\,(2 + \log N)\, t_{\text{meet}},$$

which proves the first claim.

The second claim of the proposition is [Aldous and Fill, 2002, Proposition 14.5]. □

In particular, Proposition 4.2.5 allows us to to bound the consensus time by bounding either hitting times or meeting times for an irreducible chain.

4.3 Contact Process

The contact process (or SIS infection), which we first defined in Section 2.2, can also be defined by a graphical construction. In the graphical construction of the contact process on a finite simple graph $G = ([N], E)$, we construct $2|E|$ independent Poisson processes of rate λ which contain the infection times for each directed edge. Then, i infects j at time t iff the Poisson process attached to (i, j) rings at t and also i is infected at t. Further we generate N independent Poisson processes of rate 1, corresponding to the recovery times of each vertex.

With this conception of the model, we can explicitly construct couplings between the contact process and another version with extra infection added or recovery taken away. Thus, as in [Liggett, 1999, Page 32], we can deduce various monotonicities of the contact process.

Lemma 4.3.1. *On two graphs $G^{(1)} = (V^{(1)}, E^{(1)})$, $G^{(2)} = (V^{(2)}, E^{(2)})$ with two contact processes $\xi^{(1)}$, $\xi^{(2)}$ with respective initial conditions $\mu^{(1)}$, $\mu^{(2)}$ and infection parameters $\lambda^{(1)}$, $\lambda^{(2)}$, if*

1. $V^{(1)} \subseteq V^{(2)}$
2. $E^{(1)} \subseteq E^{(2)}$
3. $\mu^{(1)} \leq \mu^{(2)}$
4. $\lambda^{(1)} \leq \lambda^{(2)}$

then we can couple the two random processes such that

$$\mathbb{P}\left(\forall t \geq 0, \forall v \in V^{(1)}, \quad \xi^{(1)}(v) \leq \xi^{(2)}(v)\right) = 1$$

i.e. we have the stochastic domination of processes $\xi^{(1)} \preceq \xi^{(2)}$.

Recall again that the contact process $(\zeta_t)_t$ has self-duality in the following form, where $\xi_t(A) := \sum_{a \in A} \xi_t(a)$.

Proposition 4.3.2 (Contact Process Self-Duality)**.** *On a static graph $G = (V, E)$ and $A, B \subset V$ the contact process satisfies, for any fixed time $t \geq 0$,*

$$\mathbb{P}\left(\xi_t(A) > 0 \middle| \xi_0 \equiv \mathbb{1}_B\right) = \mathbb{P}\left(\xi_t(B) > 0 \middle| \xi_0 \equiv \mathbb{1}_A\right).$$

When considering the recovery time of a point infection, this immediately leads to a useful equivalence between two problems. To avoid notational confusion, we adopt the following definition.

Definition 4.3.3 (Recovery time)**.** For an infection model we write $R := T_\emptyset^+$ for the hitting time of the empty infection (or the everywhere 0 infection), or the return time to this state in the case of an infection intialised at $\emptyset$ (which would not make sense for the contact process, but will for another model).

Corollary 4.3.4. *On a finite graph $G = ([N], E)$ let U be uniformly distributed in $[N]$. Write*

$$D_t := \frac{\xi_t([N])}{N}$$

for the infection density *at time t. Then we find the following relationship between contact processes initialised at $\mathbb{1}_U$ and $\mathbb{1}_{[N]}$:*

$$\mathbb{E}_U(R) = \mathbb{E}_{[N]}\left(\int_0^\infty D_t \mathrm{d}t\right).$$

Thus if we're interested in infection density as in [Mountford et al., 2013], this can in the static graph context be related back to the recovery time problem that we investigate.

Proof. By self-duality for any $v \in [N]$ we have

$$\mathbb{P}_v\left(\xi_t([N]) > 0\right) = \mathbb{P}_{[N]}\left(\xi_t(v) > 0\right) = \mathbb{E}_{[N]}\left(\xi_t(v)\right)$$

which we can sum over v to obtain

$$\mathbb{P}_U\left(\xi_t([N]) > 0\right) = \mathbb{E}_{[N]}\left(\frac{1}{N}\xi_t\right) = \mathbb{E}_{[N]}\left(D_t\right)$$

and by integration, applying Fubini's theorem

$$\mathbb{E}_U(R) = \mathbb{E}_{[N]}\left(\int_0^\infty D_t \mathrm{d}t\right).$$

□

We will apply this duality relation on the non-adaptive network in Section 7.1. Note that for the frequently considered problem of recovery time from full infection [Nam et al., 2019, Bhamidi et al., 2019], they have $A = B$ and so there is no alternative dual version of that problem.

4.3.1 Exploring with the infection

We now look at the contact process not on a static graph, but on the adaptive dynamic network model defined precisely in Definition 2.2.6. We can redefine the infection and graph dynamic from a local perspective to better understand how spreading happens.

The algorithm we describe for exploring the adaptive Erdős-Rényi graph with the contact process generalises naturally to Chung-Lu IRGs [van der Hofstad, 2016]. This

is an algorithm in the sense that each distribution identity with "$\sim$" is generated at the point of reference, independently from the past. In this algorithm we think of edges as directed from the parent to its child for the purposes of describing the local picture, but we are still actually describing an undirected graph.

At any time we have the set R_t of vertices *revealed* at time t. At time 0 we have

$$I_0 = R_0 = \{1\}, \quad \xi_0 = \mathbb{1}_1$$

so we "reveal" the vertex 1 by generating its neighbourhood i.e. giving it an outdegree

$$\mathrm{d}_0^+(1) \sim \mathrm{Bin}\left(N-1, \frac{\beta}{N}\right)$$

where each binomial trial is leading to each other unrevealed vertex with the Erdős-Rényi edge probability β/N. This is enough local information to run the contact process and the reactive graph process on this graph because a vertex cannot update without being adjacent to an infected vertex.

We *reveal* a vertex v only at the time t when it becomes infected. Its indegree is the total number of revealed vertices with an edge leading to it

$$\mathrm{d}_t^-(v) \geq 1.$$

What exactly we are revealing for this vertex is its outdegree

$$\mathrm{d}_t^+(v) \sim \mathrm{Bin}\left(N-|R_t|, \frac{\beta}{N}\right)$$

with again one trial for each remaining unrevealed vertex; note $v \in R_t$.

Furthermore at an *update* of an infected vertex v at time t (which is necessarily revealed) we delete every edge pointing into it, every edge pointing out to a revealed vertex and every "half-edge" pointing out (to an unrevealed vertex). Then we generate a new outdegree

$$\mathrm{d}_t^+(v) \sim \mathrm{Bin}\left(N-|R_t|, \frac{\beta}{N}\right),$$

a new indegree

$$\mathrm{d}_t^-(v) \sim \mathrm{Bin}\left(|R_t|-1, \frac{\beta}{N}\right),$$

and if $\mathrm{d}_t^-(v) > 0$ we identify a revealed vertex as a "parent" for each binomial trial.

Conveniently, if a vertex updates *while uninfected* it simply returns to being unrevealed and $|R_t| = |R_{t-}| - 1$; we do not generate a degree as, effectively, we have deleted it from the local picture. This vertex could further be made to update without being revealed but still there is no need to keep track of its degree.

Thus we have in a sense contained the nonstationarity – for any time $t > 0$ the edges adjacent to a vertex in R_t have a probability of presence dependent on the contact process, but other edges are independently present with probability β/N.

This means that we can use $I_t \subseteq R_t$ to say that before the epidemic event E_ϵ^N of Definition 2.2.2 each vertex when it was newly discovered had $\mathrm{d}^- \geq 1$ leading to the vertex that infected it, and edges of number

$$\mathrm{d}^+ \succeq \mathrm{Bin}\left(\lceil (1-\epsilon)N \rceil, \frac{\beta}{N}\right)$$

leading to unique unrevealed vertices that have never been infected.

By considering this description of the model, we come to a very natural approximating model which focuses on the initial phase of small $|R_t|$.

Within the first $O(1)$ vertices, a neighbourhood update will typically put the updating vertex at distance $\Theta(\log N)$ from any vertex we have as yet encountered with the infection. Therefore the approximating model that we use within the initial phase of the infection is of disjoint neighbourhoods, each being an infinite Galton-Watson tree of potentially explorable vertices.

Definition 4.3.5 (Contact Process on Evolving Forest)**.** Generate a countably infinite set of i.i.d. Galton Watson trees with offspring distribution Pois(β), labelled T^u for each label u in the Ulam-Harris tree.

The Contact Process on Evolving Forest (CPEF) has an initial tree $T^\emptyset$, with root

$$v_\emptyset^\emptyset \in T^\emptyset$$

and at time 0 this is the only infected vertex. Its children are denoted $v_1^\emptyset$, $v_2^\emptyset$ and so on in the Ulam-Harris convention. On this structure, we put the contact process dynamic: that infected vertices infect each of their neighbours independently at rate λ, while recovering at rate 1.

Further, as in the network model, each vertex "updates" at rate κ if and only if it has an infected neighbour. We number the vertex $v_\emptyset^\emptyset$ with 1 and continue numbering the other vertices in the lexicographical (breadth first) ordering. Thus when a vertex at lexicographical order k updates in $T^\emptyset$, it is deleted in that graph and identified with the

root of T^k, so that it will infect the root of T^k at that point in time if it is currently infected. Then, vertices in T^k are also numbered by the lexicographical ordering so that if some vertex in position m in that tree does update, it is deleted and identified with the root of $T^{k,m}$.

Thus any "tree" set T^u may in fact span a disconnected graph in the course of the process, but it will never contain a cycle. We write I^u_∞ for the vertices in T^u which ever become infected.

We can think of this CPEF version of the dynamics as being obtained from the dynamics on the network by taking $N \to \infty$ in the local weak limit.

4.3.2 The subtree contact process

Despite how simple the contact process is to describe, it is hard to analyse directly – and the self-duality cannot help us move to a more tractable model. Hence, we analyse is by constructing stochastic bounds: for example we see the following upper bound model in [Liggett, 1999].

Definition 4.3.6 (Branching random walk infection)**.** The branching random walk infection on a graph $G = (V, E)$ is obtained from the contact process by *neglecting exclusion.* This means that in the branching random walk infection $\zeta_t \in \mathbb{N}^V$, each infected agent can infect every one of its neighbours, and in the event that they are already infected they will be *multiply infected.* $\zeta_t(v) = k$ reflects that k infections occupying v will all independently infect $\Gamma(v)$ – or equivalently, each neighbour is infected at rate λk.

By the graphical construction we described, it is clear that this process has strictly more infected particles and so ζ_t is indeed a stochastic upper bound on the contact process. Analysis is done in this case by constructing supermartingales. On the d-regular tree T_d take an arbitrary root $\emptyset \in V(T_d)$ and let $l(v) := d(\emptyset, v)$ record the distance from the root. For some $\rho > 0$, these supermartingales then have the simple form

$$M_t = \sum_{v \in V(T_d)} \zeta_t(v) \rho^{l(v)}$$

so that by finding conditions where the $\rho = 1$ process is supermartingale we can prove global extinction, or by the same but setting $\rho < 1$ we can show local extinction in the neighbourhood of $\emptyset$. In fact in the case $\rho = 1$, M_t is simply a continuous time Galton-Watson process. We can then observe that each action of a particle is independently

either an infection or recovery, and is an infection with probability

$$\frac{\lambda d}{1+\lambda d}$$

and so by comparison to the biased random walk the following result is straightforward.

Theorem 4.3.7. *Consider the branching random walk infection $(\zeta_t)_t$ on $T_d \ni \emptyset$. The expected recovery time of a branching random walk infection from $\emptyset$ satisfies*

$$\mathbb{E}\left(\inf\{t : \zeta_t \equiv 0\} \,\middle|\, \zeta_0 = \mathbb{1}_\emptyset\right) \begin{cases} < \infty & \lambda < \frac{1}{d}, \\ = \infty & \lambda \geq \frac{1}{d}. \end{cases}$$

We know from [Pemantle, 1992, Theorem 2.2] that the real contact process cannot survive on T_d unless $\lambda \geq \frac{1}{d-1}$, so naturally the branching random walk infection did miss a region of subcritical parameters. We will construct a different upper bound which is tractable not by any associated supermartingale but by being a positive recurrent, reversible Markov chain. However, to achieve such a nice stochastic bound model, we will see by comparison to the above theorem that we sacrifice even more viable λ parameters than we did in using the branching random walk infection, with a process which is similar to the Root-Added Contact Process of [Bhamidi et al., 2019].

Definition 4.3.8 (Subtree Contact Process)**.** Construct the Subtree Contact Process (SCP) on a rooted tree $(T, \emptyset)$ by adding an additional vertex adjacent to the root $\emptyset$. This additional vertex is infected and cannot recover and thus the contact process has no absorbing state. We further modify the contact process so that no other vertex can recover unless all of its children are healthy.

This defines the dynamic, on the set $\mathscr{T}$ of finite subtrees of T which are rooted at $\emptyset$ (the extra permanently infected vertex is not considered part of the infection set).

Further, if $Z_\lambda := \sum_{T \in \mathscr{T}} \lambda^{|T|} < \infty$, write for $T \in \mathscr{T}$

$$\pi_\lambda(T) := \frac{\lambda^{|T|}}{Z_\lambda} \tag{4.12}$$

for the SCP stationary distribution. Note that the SCP defined thus is irreducible and *reversible* on this set of subtrees, and we can check the stationary distribution (4.12) in the detailed balance equations.

Because we made both simplifications by adding infection or preventing recovery, the SCP infection stochastically dominates the usual contact process. We will therefore use

the SCP to contain the infection sets.

Because we have constucted an upper bound which is a reversible Markov chain, we can get a good understanding, for example, of the *hitting probabilities*. Define for a rooted subtree t

$$p(t) := \mathbb{P}_{\emptyset}^{\lambda}\left(T_t < T_{\emptyset}^{+}\right)$$

then given $|t| = k$ we can give fairly good a.s. bounds on $p(t)$ via the electrical network techniques.

Proposition 4.3.9. *If the SCP lives on a tree such that $Z_\lambda < \infty$, we find*

$$(1-\lambda)\lambda^{k-1} \leq p(t) \leq k\lambda^{k-1}$$

for a rooted tree t with $|t| = k$ vertices.

Proof. Let $\mathbf{E}$ be the set of "edges" corresponding to a recovery at t. We calculate the total conductance of this set (see (4.1)) from the number $\ell(t)$ of leaves

$$c(\mathbf{E}) = \ell(t)\frac{\lambda^{|t|}}{Z_\lambda} \leq \frac{k\lambda^k}{Z_\lambda}$$

so that by the Nash-Williams inequality [Levin et al., 2009, Proposition 9.15] we deduce

$$\mathcal{R}(\emptyset \leftrightarrow t) \geq \frac{1}{c(\mathbf{E})} \geq \frac{Z_\lambda}{k\lambda^k}.$$

Further, by Thompson's Principle [Levin et al., 2009, Theorem 9.10] applied to an arbitrary simple path from $\emptyset$ to t of minimal length k, we find

$$\mathcal{R}(\emptyset \leftrightarrow t) \leq \sum_{s=1}^{k} \frac{Z_\lambda}{\lambda^s} = \frac{\lambda^{-k}-1}{1-\lambda} Z_\lambda \leq \frac{\lambda^{-k}}{1-\lambda} Z_\lambda.$$

These two bounds can be combined with the relation

$$p(t) = \frac{1}{c(\emptyset)\mathcal{R}(\emptyset \leftrightarrow t)} = \frac{Z_\lambda}{\lambda \mathcal{R}(\emptyset \leftrightarrow t)}$$

to deduce

$$1 - \lambda \leq \frac{p(t)}{\lambda^{k-1}} \leq k.$$

□

As warned, the use of this upper bound does require smaller λ parameters than the

branching random walk infection. We see this by a comparison on the d-regular tree.

Theorem 4.3.10. *Let $(\zeta_t)_t$ denote the SCP on T_d with arbitrary root $\circledcirc$, from initial distribution $\zeta_0 = \{\circledcirc\} =: r$. For this model we find threshold*

$$\lambda_c = \frac{(d-2)^{d-2}}{(d-1)^{d-1}}$$

which has recovery time (of Definition 4.3.3)

$$\mathbb{E}\left(R\middle|\zeta_0 = r\right)\begin{cases} < \infty & \lambda < \lambda_c \\ = \infty & \lambda > \lambda_c. \end{cases}$$

To make the comparison, we further observe that $\lambda_c \sim \frac{1}{ed}$ as $d \to \infty$.

Proof. See Proposition A.0.1 for a proof that, in $\mathscr{T}$ the set of rooted subtrees of the rooted d-regular tree T_d, we have

$$\frac{d}{k-1}\binom{k(d-1)}{k-2} =: C_k$$

rooted subtrees of size k. By Stirling's approximation as $k \to \infty$ we observe

$$\log C_k = (d-1)k\log((d-1)k) - (d-1)k - k\log k + k - (d-2)k\log(d-2)k + (d-2)k + o(k)$$

and so after cancellation we can deduce the limit

$$\frac{\log C_k}{k} \to (d-1)\log(d-1) - (d-2)\log(d-2).$$

Hence if we define

$$\lambda_c := \frac{(d-2)^{d-2}}{(d-1)^{d-1}}$$

the putative normalising constant Z_λ has

$$Z_\lambda = \sum_{T\in\mathscr{T}} \lambda^{|T|} = 1 + \sum_{k=1}^{\infty} C_k\lambda^k \begin{cases} < \infty & \text{if } \lambda < \lambda_c \\ = \infty & \text{if } \lambda > \lambda_c. \end{cases}$$

so that we have a stationary distribution $\pi_\lambda(T) \propto \lambda^{|T|}$ if $\lambda < \lambda_c$. By considering a return

time from $\emptyset$, which must first feature an infection of the root tree r, we deduce

$$\mathbb{E}_r(R) = \frac{1}{\lambda \pi_\lambda(\emptyset)} - \frac{1}{\lambda} = \frac{Z_\lambda - 1}{\lambda} < \infty.$$

In the other case, nonexistence of a stationary distribution and irreducibility of the chain implies by [Norris, 1998, Theorem 3.5.3] that the expected return times are infinite.

Now

$$d \cdot \lambda_c \sim \left(\frac{d-2}{d-1}\right)^{d-1} = \left(1 - \frac{1}{d-1}\right)^{d-1} \to \frac{1}{e}$$

which proves the final claim. □

We will reguire two monotonicities of the SCP for its analysis. With the graphical construction these are reasonably intuitive but for the sake of completeness we will prove them here.

Proposition 4.3.11. *Let $\zeta^{(1)}$ be the SCP with infection rate $\lambda > 0$ on the tree $T^{(1)}$ with root $\circledcirc$, and $\zeta^{(2)}$ be the SCP with infection rate λ on the tree $T^{(2)}$ which contains $T^{(1)}$ and has the same root $\circledcirc$. Then the two processes started from $\zeta_0^{(1)} = \zeta_0^{(2)} = \{\circledcirc\}$ have the process domination $\zeta^{(1)} \preceq \zeta^{(2)}$.*

Proof. Note that $\zeta_0^{(1)} \subseteq \zeta_0^{(2)}$ which will be the base case for the induction.

Construct $\zeta^{(2)}$ by the graphical construction, attaching a Poisson process of rate λ to each directed edge containing the possible infection times and a Poisson process of rate 1 to each vertex containing the possible recovery times. Then we have immediately a coupled construction for $\zeta^{(1)}$ by using the same Poisson processes whenever they are attached to a vertex or directed edge in $T^{(1)}$.

Now we check that $\zeta_t^{(1)} \subseteq \zeta_t^{(2)}$ is preserved by the dynamics. If $\zeta^{(2)}$ has a successful recovery event in $v \in V(T^{(1)})$ at time t then $\zeta^{(1)}$ must also attempt to recover there at that time. Hence either $v \notin \zeta_{t-}^{(1)}$ or $\zeta^{(1)}$ also has a successful recovery event in v at time t, which in either case preserves the inclusion.

We must also consider the event that $v \in \zeta_{t-}^{(1)}$ infects $w \in V(T^{(1)})$ at time t. By our inductive hypothesis, $v \in \zeta_{t-}^{(2)}$. Hence either $w \in \zeta_{t-}^{(2)}$ or we see the same infection in $\zeta^{(2)}$, because the modifications to the contact process to define the SCP in Definition 4.3.8 did not affect the infection events.

We do not need to consider the vertices in $V(T^{(2)}) \setminus V(T^{(1)})$ because $\zeta^{(1)}$ can never include these vertices, and so this concludes the induction. □

Proposition 4.3.12. *Let $\zeta^{(1)}$ be the SCP with infection rate $\lambda_1 > 0$ on the tree T with root o, and $\zeta^{(2)}$ be the SCP with infection rate $\lambda_2 > \lambda_1$ on the same rooted tree T. Then the two processes with $\zeta_0^{(1)} = \zeta_0^{(2)} = \{o\}$ have the process domination $\zeta^{(1)} \preceq \zeta^{(2)}$.*

Proof. This is proved in much the same way and the argument in checking the recovery events for $\zeta^{(2)}$ is identical.

The coupling for the infection events is constructed by first giving $\zeta^{(1)}$ a graphical construction with infection rates λ_1. Then to each of these processes we add an independent Poisson process of rate $\lambda_2 - \lambda_1$ which creates a graphical construction with infection rates λ_2. Thus any infection attempt in $\zeta^{(1)}$ also exists for $\zeta^{(2)}$ and the inclusion is also preserved here. □

Recent work of [Nam et al., 2019] managed to find an extinction-survival threshold in the more difficult context of a Poisson(β)-Galton-Watson tree by taking the limit $\beta \to \infty$. This was established in the sense that the random survival threshold

$$\lambda_1(\beta) := \inf\{\lambda > 0 : \text{positive probability to survive from full infection}\}$$

decays like $\lambda_1(\beta) \sim \frac{1}{\beta}$ as $\beta \to \infty$. We can quickly prove a similar claim for the SCP by looking at when the Poisson-Galton-Watson tree contains the regular tree.

Theorem 4.3.13 ([Pakes and Dekking, 1991])**.** *There exist β such that the rooted* Pois(β)*-Galton-Watson tree has positive probability of a (rooted) T_d subtree that shares its root. This occurs in fact for all $\beta > \beta_d$ where the minimal such value β_d has*

$$\beta_{d+1} \sim d$$

as $d \to \infty$.

We also state a result that will be proved in Chapter 7 but is useful for this discussion.

Lemma 4.3.14. *On the random (finite or infinite)* Pois(β)*-Galton-Watson tree we have $\mathbb{E}(Z_\lambda) < \infty$ whenever $\lambda < \frac{1}{\beta e}$.*

Recall, for this corollary, that $r := \{\emptyset\}$ is the subtree containing only the root.

Corollary 4.3.15. *Fix a constant $C > 0$ and consider infection rate*

$$\lambda_\beta := \frac{C}{\beta}.$$

We find for the SCP on a rooted Pois(β)*-Galton-Watson tree that*

$$\mathbb{E}_r(R) \begin{cases} < \infty & \text{if } C < \frac{1}{e} \\ = \infty & \text{if } C > \frac{1}{e} \end{cases}$$

for β sufficiently large.

Proof. First if $C < \frac{1}{e}$ we find

$$\lambda = \lambda_\beta = \frac{C}{\beta} < \frac{1}{\beta e}$$

and conclude immediately from Lemma 4.3.14.

If instead $C > \frac{1}{e}$, consider the maximal value k_β such that the regular tree T_{k_β} has positive probability to be contained in the rooted Pois(β)-Galton-Watson tree and include the root of the Galton-Watson tree. By Theorem 4.3.13 we have $k_\beta \sim \beta$ as $\beta \to \infty$.

Then recall from Theorem 4.3.10 that the SCP with parameter $\lambda = \lambda_\beta$ will have infinite expected recovery time on T_{k_β} if

$$\frac{C}{\beta} > \frac{1}{ek_\beta} \tag{4.13}$$

and k_β is sufficiently large. But observe that (4.13) holds for β sufficiently large by the assumption $C > \frac{1}{e}$. Hence by the subtree monotonicity of Proposition 4.3.11 we infer that on the event that T_{k_β} is a subtree, recovery happens strictly later. Because recovery time on T_{k_β} has infinite expectation, we conclude the same for the Galton-Watson tree. □

Interestingly, this positive recurrence region $\lambda\beta < \frac{1}{e}$ is the same as for the MRACP of [Bhamidi et al., 2019]. The MRACP only disallows recovery at the root while its children are infected rather than at every vertex as for the SCP – while this sounds like a bigger modification, it seems we have not lost ground to any combination effects.

Chapter 5

Voter Model Consensus on a Subcritical Scale-Free Network

In this chapter we will prove the two main theorems about the asymptotics of the consensus time on a subcritical network in the class $\mathscr{G}_{\beta,\gamma}$ of Definition 2.0.5. In Section 5.1, we will consider the classical voter model and prove Theorem 2.1.5 on the change in the polynomial order of the expected consensus time with $\theta \in \mathbb{R}$. Then, in Section 5.2 we will prove the corresponding result Theorem 2.1.7 for the discursive voter model. Throughout, we will use the duality of the voter model to a system of coalescing random walks as described in Section 4.2. Appropriate tools to bound the expected coalescence times are combined with a fine analysis of the structure of subcritical inhomogeneous random graph models, which have the fractal-like structure seen in Figure 5.1, that is not readily available in the literature.

A type of mean-field approximation has also been used to analyse the voter model in [Sood et al., 2008] where we believe it was very successful. However, they do not get the right polynomial order of consensus time in the case of subcritical network parameters, because the lack of a giant component means that treating the graph as connected no longer works. Therefore we will focus on the subcritical case in this chapter and in doing so we find a complete picture of new polynomial orders.

Proving these two theorems boils down to sixteen bounds: a two sided bound for each of the four regimes in each theorem. Of these bounds the more illustrative are the lower bounds as these point to a component which is creating the slow consensus time over the network. Technically, these lower bounds are either by the mixing lower bound

of Proposition 4.1.4 (b)

$$\max_{A\subset[N]} \frac{\pi(A)\pi(A^c)}{\sum_{x\in A}\sum_{y\in A^c} c(xy)}$$

or by the other lower bound of Proposition 4.1.4 (a), always on the order

$$\frac{1}{\sum_{i\in[N]} q(i)\pi(i)^2}.$$

This second bound we think of as summing up all the effects apart from mixing: the fastest mixing approximation of the chain should be that with a jump chain distribution that is simply a sequence of stationary draws. If we consider this chain while preserving the jump rates $q(\cdot)$, we would find a consensus time of order $\frac{1}{\sum q\pi^2}$. So we can say that if this is the true consensus time for the model then it would not be accelerated by moving to the fastest mixing version of the consensus dynamic and so, in this sense, is not bottlenecked by mixing.

The exponents that we see in the classical voter model on G_N, which we recall are

$$\mathbb{c} = \begin{cases} \gamma & \theta \geq 1, \\ \gamma\theta & \frac{1}{2-2\gamma} < \theta < 1, \\ \frac{\gamma}{2-2\gamma} & 0 \leq \theta \leq \frac{1}{2-2\gamma}, \\ \frac{\gamma(1-\theta)}{2-2\gamma} & \theta < 0, \end{cases}$$

can be divided into two regimes. If $\theta < \frac{1}{2-2\gamma}$ the bottleneck is by slow *mixing*, on a component of polynomial order $N^{\frac{\gamma}{2-2\gamma}}$. In the other case $\theta > \frac{1}{2-2\gamma}$ we have sufficiently fast mixing and the slowest consensus is instead seen on the largest components.

For the discursive voter model we instead found

$$\mathbf{c} = \begin{cases} \frac{\gamma}{2-2\gamma} & \theta \geq \frac{3-4\gamma}{2-2\gamma}, \\ \gamma(2-\theta) & 1 < \theta < \frac{3-4\gamma}{2-2\gamma}, \\ \gamma & 2\gamma \leq \theta \leq 1, \\ \frac{\gamma(2-\theta)}{2-2\gamma} & \theta < 2\gamma. \end{cases}$$

and the mixing bottlenecks on components of order $N^{\frac{\gamma}{2-2\gamma}}$ are seen in the edge regions $\theta \in (-\infty, 2\gamma) \cup \left(\frac{3-4\gamma}{2-2\gamma}, \infty\right)$. One difference here is that we cannot demonstrate these two mixing lower bounds on the same components – it will be necessary to construct a double star and also an elongated version of the double star.

	Classical		Discursive	
	Upper bound	Lower bound	Upper bound	Lower bound
Very high θ	Hitting time 4.2.5	Mixed meeting 4.1.4 (a)	Meeting in leaves 4.1.10	Mixing 4.1.4 (b)
High θ	Meeting in leaves 4.1.10	Mixed meeting 4.1.4 (a)	Meeting in leaves 4.1.10	Mixed meeting 4.1.4 (a)
Low θ	Meeting in leaves 4.1.10	Mixing 4.1.4 (b)	Meeting in leaves 4.1.10	Mixed meeting 4.1.4 (a)
Very low θ	Meeting at centre 4.1.7	Mixing 4.1.4 (b)	Meeting in leaves 4.1.10	Mixing 4.1.4 (b)

Table 5.1: A summary of the methods by which we obtain each of the 16 required bounds in the proofs of Theorems 2.1.5 and 2.1.7. With each named method we refer to the proposition or theorem which formalises the bounding concept, and note also that the "leaves" referred to are only the leaves adjacent to the central vertex as in Proposition 3.2.3 (b).

5.1 Consensus Time for the Classical Voter Model

In this section, we will consider the classical voter model as defined in Definition 2.1.3. Throughout, let the SNR model G_N of Definition 2.0.4 with for $\beta + 2\gamma < 1$ be the underlying graph, but recall that results apply across the family of subcritical rank-one networks related to G_N by condition (2.2). We note that this version of the voter model fits into the general setting of a Q-voter model of Section 2.1 if for $\theta \in \mathbb{R}$ we consider $Q = Q^\theta$ defined as

$$Q^\theta(i,j) = \mathrm{d}(i)^{\theta-1} \quad \text{if } i \sim j \text{ in } G_N. \tag{5.1}$$

As before, we write $\mathbb{P}^\theta$ for the law of (and $\mathbb{E}^\theta$ for the expectation with respect to) the coalescing random walks with generator Q^θ.

If we denote by $\mathscr{C}_1, \dots, \mathscr{C}_k$ the connected components of G_N, then these also correspond to the irreducible components of the Markov chain with generator Q^θ. So if we let $\pi = (\pi(z), z \in V(G_N))$ be defined via

$$\pi(z) = \frac{\mathrm{d}(z)^{1-\theta}}{\sum_{y \in \mathscr{C}_j} \mathrm{d}(y)^{1-\theta}}, \qquad \text{for } z \in \mathscr{C}_j,$$

for $j \in [k]$ (or naturally $\pi(z) = 1$ if $\mathrm{d}(z) = 0$), then $\pi|_{\mathscr{C}_j}$ is the invariant measure of the Q^θ Markov chain restricted to $\mathscr{C}_j$.

Before the main proof, we show an elementary bound on the meeting time of two independent random walks, when the component contains a star, i.e. if there exists a vertex k with a set of neighbours L_k, each of degree 1 (compare Proposition 3.2.3).

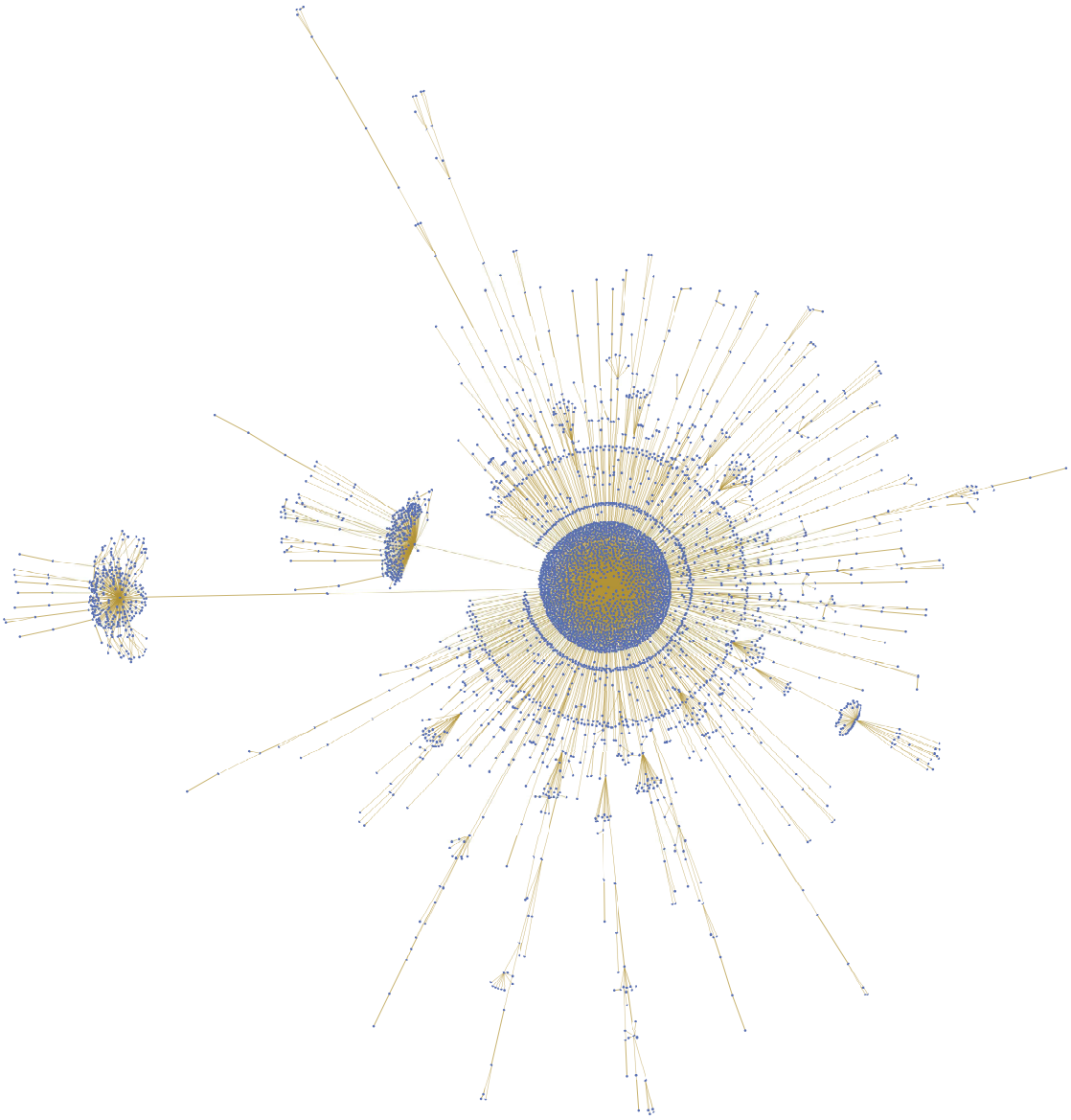


Figure 5.1: The component containing the vertex 1, for a graph in the class $\mathcal{G}_{\beta,\gamma}$ with subcritical network parameters $(\beta, \gamma) = (0.05, 0.45)$. On these 4429 vertices we can already see the emerging fractal structure.

Lemma 5.1.1. *Let $k \in [N]$ be such that L_k, the set of its neighbours of degree 1, is non empty. Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be independent Markov chains on $\mathscr{C}(k)$ with generator Q^θ. Then, for the product chain observed on $\{k\} \cup L_k$ (as defined before Theorem 4.1.10), we have*

$$t^\pi_{\text{meet}}(\{k\} \cup L_k) \leq \frac{3 + \mathrm{d}(k)^\theta}{2}.$$

Proof. Let S_t count how many of the two walkers are currently in the leaf set L_k. Then

$(S_t)_{t\geq 0}$ is a Markov chain on $\{0,1,2\}$ with transition rates (s_{ij}), where in particular

$$s_{21} = 2, \quad s_{10} \geq 1, \quad s_{12} = |L_k|\,\mathrm{d}(k)^{\theta-1} \leq \mathrm{d}(k)^{\theta},$$

using that $|L_k| \leq \mathrm{d}(k)$.

Now, note that $S_t = 0$ implies that $\tau_{\mathrm{meet}} \leq t$. In particular, if $T_i = \inf\{t \geq 0 : S_t = i\}$ for $i \in \{0,1,2\}$, then we have that $T_0 \geq \tau_{\mathrm{meet}}$.

From the explicit transition rates, we can see that $\mathbb{E}_2(T_1) = \frac{1}{2}$ and if we write $s(1) = s_{10} + s_{12}$, then

$$\mathbb{E}_1(T_0) = \frac{1}{s(1)} + \frac{s_{12}}{s(1)}\left(\frac{1}{2} + \mathbb{E}_1(T_0)\right)$$

so that

$$\mathbb{E}_1(T_0) = \frac{1 + s_{12}/2}{s_{10}} \leq 1 + \frac{s_{12}}{2}.$$

We conclude that

$$\sup_{v,w\in V(\mathcal{S}_k)} \mathbb{E}_{(v,w)}(\tau_{\mathrm{meet}}) \leq \max\{\mathbb{E}_1(T_0), \mathbb{E}_2(T_0)\} \leq \frac{1}{2} + 1 + \frac{s_{12}}{2} \leq \frac{3 + \mathrm{d}(k)^{\theta}}{2},$$

as claimed. □

Proof of Theorem 2.1.5. We will start by showing the **upper bounds**. For the cases $\theta \geq 1$, we use that by Lemma 4.2.4 and Proposition 4.2.5 we can bound

$$\mathbb{E}_{\mu_u}(\tau_{\mathrm{cons}} \,|\, G_N) \leq e(2 + \log N) t_{\mathrm{hit}}(G_N),$$

where $t_{\mathrm{hit}}(G_N) = \sup_{j\in[k]} t_{\mathrm{hit}}(\mathscr{C}_i)$ for $\mathscr{C}_1, \ldots, \mathscr{C}_k$ the components of G_N. Note in particular that the right hand side is still random and the expectation is only over the random walks.

We recall that the random walk associated to the classical voter model has transition rates $Q^{\theta}(x,y) = \mathrm{d}(x)^{\theta-1}\mathbb{1}_{x\sim y}$ and $\pi(x) \propto \mathrm{d}(x)^{1-\theta}$. In particular, for any component $\mathscr{C} \in \mathrm{Comp}(G_N)$ the conductances as defined in (4.1) are

$$c(xy) = \pi(x)Q^{\theta}(x,y) = \frac{1}{\sum_{z\in\mathscr{C}} \mathrm{d}(z)^{1-\theta}} \mathbb{1}_{\{x\sim y\}}, \quad \text{for any } x, y \in \mathscr{C}.$$

Hence, by Proposition 4.1.1, we have for any component $\mathscr{C}$,

$$t_{\text{hit}}(\mathscr{C}) \leq \text{diam}(\mathscr{C}) \sum_{z \in \mathscr{C}} \text{d}(z)^{1-\theta}. \tag{5.2}$$

Because $\theta \geq 1$, we have that $\sum_{z \in \mathscr{C}} \text{d}(z)^{1-\theta} \leq |\mathscr{C}|$. Therefore, by Proposition 3.2.2 (a) and Proposition 3.2.1, we get that

$$\sup_{\mathscr{C} \in \text{Comp}(G_N)} t_{\text{hit}}(\mathscr{C}) \leq \max_{\mathscr{C} \in \text{Comp}(G_N)} \text{diam}(\mathscr{C}) \times \max_{\mathscr{C} \in \text{Comp}(G_N)} |\mathscr{C}| = O_{\mathbb{P}}^{\log N}(N^{\gamma}),$$

which completes the upper bound for $\theta \geq 1$.

For $\theta \leq 0$, we first deal with the small components, where we recall that the vertex set of the 'small' components is defined as

$$V_{\text{small}} := [N] \setminus V_{\text{big}}, \qquad \text{where } V_{\text{big}} := \bigcup_{k \leq K_{\gamma}} V\left(\mathscr{C}(k)\right)$$

and $K_{\gamma} = N^{\frac{1-2\gamma}{2-2\gamma}} \log N$. By Proposition 3.2.2 (b), we know that

$$\max_{k \in V_{\text{small}}} \sum_{x \in \mathscr{C}(k)} \text{d}(x) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$$

In particular, we get from (5.2) using $\sum_i x_i^p \leq (\sum_i x_i)^p$ for any $p \geq 1$ and $x_i \geq 0$ that

$$\begin{aligned} \max_{k \in V_{\text{small}}} t_{\text{hit}}(\mathscr{C}(k)) &\leq \text{diam}(G_N) \max_{k \in V_{\text{small}}} \sum_{x \in \mathscr{C}(k)} \text{d}(x)^{1-\theta} \\ &\leq \text{diam}(G_N) \max_{k \in V_{\text{small}}} \Big(\sum_{x \in \mathscr{C}(k)} \text{d}(x) \Big)^{1-\theta} = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(1-\theta)}{2-2\gamma}}\right), \end{aligned} \tag{5.3}$$

where we also used Proposition 3.2.1 to bound the diameter.

To bound the consensus time on large components, we use that by Proposition 4.1.7 for any $k \leq K_{\gamma}$,

$$t_{\text{meet}}(\mathscr{C}(k)) \leq 189 \, \frac{t_{\text{hit}}(k)}{\pi(k)},$$

and find a suitable upper bound on the right hand side, which in turn gives us by Lemma 4.2.4 and Proposition 4.2.5 an upper bound on $\mathbb{E}_{\mu_u}(\tau_{\text{cons}}(\mathcal{C}(k)) \,|\, G_N)$. In order to bound the invariant measure, we note that since $\theta \leq 0$, we have from Proposition

3.2.2 (a) and 3.2.3 (a) that

$$\min_{k\le K_\gamma} \pi(k) = \min_{k\le K_\gamma} \frac{\mathrm{d}(k)^{1-\theta}}{\sum_{z\in\mathscr{C}(k)} \mathrm{d}(v)^{1-\theta}} \ge \min_{k\le K_\gamma} \Big(\frac{\mathrm{d}(k)}{\sum_{z\in\mathscr{C}(k)} \mathrm{d}(z)}\Big)^{1-\theta} = \Omega_{\mathbb{P}}^{\log N}(1).$$

In order to bound the hitting time $t_{\mathrm{hit}}(k)$ we apply the same argument as for the small component, but for the random walk restricted to each branch of $\mathscr{C}(k)$ (see also Definition 3.2.4 above for the formal definition of a branch). The bound on the sum of degrees comes from Lemma 3.2.5. Together, we obtain that

$$\sup_{k\le K_\gamma} t_{\mathrm{meet}}(\mathscr{C}(k)) = O_{\mathbb{P}}^{\log N}\Big(N^{\frac{\gamma(1-\theta)}{2-2\gamma}}\Big).$$

Combined with the bound on the small components, this completes the upper bound in the case $\theta \le 0$.

We complete the upper bounds by showing for $\theta \in (0,1)$ that

$$\max_{k\in[N]} t_{\mathrm{meet}}\left(\mathscr{C}(k)\right) = O_{\mathbb{P}}^{\log N}\left(N^{\gamma\theta} + N^{\frac{\gamma}{2-2\gamma}}\right). \tag{5.4}$$

The upper bound on the consensus time then follows by Proposition 4.2.5 and by noting that in each of the two different regimes one of the summands dominates.

For $k \in V_{\mathrm{small}}$, we use similar strategy as above and obtain by (5.2) that

$$t_{\mathrm{hit}}(\mathscr{C}(k)) \le \mathrm{diam}(\mathscr{C}(k)) \sum_{z\in\mathscr{C}(k)} d(z)^{1-\theta} \le \mathrm{diam}(\mathscr{C}(k)) \sum_{z\in\mathscr{C}(k)} d(z),$$

which if we combine Proposition 3.2.2 (b) and Proposition 3.2.1 is seen to be $O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$ uniformly in $k \in V_{\mathrm{small}}$.

For the bound on the large components, define for $k \le K_\gamma$ the set L_k as the neighbours of k that have degree 1. By Proposition 3.2.3 (b) we have that

$$\min_{k\le K_\gamma} \frac{|L_k|}{\mathrm{d}(k)} = \Omega_{\mathbb{P}}(1). \tag{5.5}$$

Hence, since all vertices in L_k have degree 1, we obtain

$$\pi(L_k \cup \{k\}) \ge \frac{\sum_{x\in L_k} \mathrm{d}(x)^{1-\theta}}{\sum_{x\in\mathscr{C}(k)} \mathrm{d}(x)^{1-\theta}} \ge \frac{|L_k|}{\sum_{x\in\mathscr{C}(k)} \mathrm{d}(x)}.$$

Thus by (5.5) and Proposition 3.2.2 (a) we have

$$\min_{k\le K_\gamma} \pi(L_k \cup \{k\}) = \Omega\left(\min_{k\le K_\gamma} \frac{\mathrm{d}(k)}{\sum_{x\in\mathscr{C}(k)} \mathrm{d}(x)}\right) = \Omega_{\mathbb{P}}\left(\frac{1}{\log N}\right). \tag{5.6}$$

Because we have a large stationary mass in $L_k \cup \{k\}$, Theorem 4.1.10 gives us that

$$t_{\mathrm{meet}}(\mathscr{C}(k)) = O_{\mathbb{P}}^{\log N}\left(t^{\pi}_{\mathrm{meet}}(L_k \cup \{k\}) + t_{\mathrm{hit}}(k)\right), \tag{5.7}$$

where we recall that $t^{\pi}_{\mathrm{meet}}(L_k \cup \{k\})$ is the meeting time for the Markov chain observed on $\{k\} \cup L_k$ (see also the definition just before Theorem 4.1.10). We obtain from by Lemma 5.1.1 that

$$\max_{k\le K_\gamma} t^{\pi}_{\mathrm{meet}}(\{k\} \cup L_k) \le \max_{k\le K_\gamma} \frac{3 + \mathrm{d}(k)^\theta}{2} = O_{\mathbb{P}}(N^{\gamma\theta}).$$

Moreover, by Lemma 3.2.5

$$\begin{aligned}\max_{k\le K_\gamma} t_{\mathrm{hit}}(k) &\le \max_{k\le K_\gamma} \max_{B\in\mathcal{B}(\mathscr{C}(k))} \mathrm{diam}(B) \sum_{v\in B} \mathrm{d}(v)^{1-\theta} \\ &\le \max_{k\le K_\gamma} \mathrm{diam}(G_N) \max_{B\in\mathcal{B}(\mathscr{C}(k))} \sum_{v\in B} \mathrm{d}(v) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right).\end{aligned} \tag{5.8}$$

Substituting both bounds into (5.7), we obtain

$$\max_{k\le K_\gamma} t_{\mathrm{meet}}(\mathscr{C}(k)) = O_{\mathbb{P}}^{\log N}\left(N^{\gamma\theta} + N^{\frac{\gamma}{2-2\gamma}}\right).$$

By combining this with the bound on the small components, we have completed the proofs for the upper bounds in all cases.

We continue with the **lower bounds**.

For the first part, we suppose that $\theta > 0$ and consider the consensus time on $\mathscr{C}(1)$. By Lemma 4.2.4 and Proposition 4.2.5

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\mathrm{cons}}(\mathscr{C}(1)) \,|\, G_N) \ge 2u(1-u)t_{\mathrm{meet}}(\mathscr{C}(1)) \ge 2u(1-u)t^{\pi}_{\mathrm{meet}}(\mathscr{C}(1)),$$

where the last inequality follows from the definitions. To bound the right hand side, we recall from Proposition 4.1.4 (a) that

$$t^{\pi}_{\mathrm{meet}}(\mathscr{C}(1)) \ge \frac{(1 - \sum_{x\in\mathscr{C}(1)} \pi(x)^2)^2}{4\sum_{x\in\mathscr{C}(1)} q(x)\pi(x)^2}. \tag{5.9}$$

In order to find a lower bound on the right hand side, we first bound the maximum of the invariant distribution. If $\theta \in (0,1)$, then we have by Proposition 3.2.3 (a)

$$\max_{v\in\mathscr{C}(1)} \pi(v) = \frac{\max_{v\in\mathscr{C}(1)} \mathrm{d}(v)^{1-\theta}}{\sum_{z\in\mathscr{C}(1)} \mathrm{d}(z)^{1-\theta}} \le \frac{|\mathscr{C}(1)|^{1-\theta}}{|\mathscr{C}(1)|} \le \mathrm{d}(1)^{-\theta} = O_{\mathbb{P}}(N^{-\gamma\theta}).$$

Similarly, if $\theta \ge 1$ we recall the leaf neighbours of Proposition 3.2.3 (b),

$$\max_{v\in\mathscr{C}(1)} \pi(v) \le \frac{1}{\sum_{z\in\mathscr{C}(1)} \mathrm{d}(z)^{1-\theta}} \le \frac{1}{|L_1|} = O_{\mathbb{P}}(N^{-\gamma}).$$

In particular, in both cases we have

$$\sum_{x\in\mathscr{C}(1)} \pi(x)^2 \le \max_{x\in\mathscr{C}(1)} \pi(x) \sum_{v\in\mathscr{C}(1)} \pi(v) = o_{\mathbb{P}}(1).$$

To estimate the denominator in (5.9), we note that for $\theta \ge 1$,

$$\begin{aligned}\sum_{v\in\mathscr{C}(1)} q(v)\pi(v)^2 &= \frac{\sum_{v\in\mathscr{C}(1)} \mathrm{d}(v)^{\theta}\, \mathrm{d}(v)^{2-2\theta}}{\Big(\sum_{v\in\mathscr{C}(1)} \mathrm{d}(v)^{1-\theta}\Big)^2} \le \frac{\sum_{v\in\mathscr{C}(1)} \mathrm{d}(v)}{|L_1|^2} \\ &= O_{\mathbb{P}}^{\log N}\Big(\frac{N^{\gamma}}{N^{2\gamma}}\Big) = O_{\mathbb{P}}^{\log N}(N^{-\gamma}),\end{aligned}$$

where we used Proposition 3.2.3 (b) for the denominator and Proposition 3.2.2 (a) for the numerator. By the same results and Lemma 3.2.6, we have for $\theta \in (0,1)$,

$$\sum_{v\in\mathscr{C}(1)} q(v)\pi(v)^2 \le \frac{\sum_{v\in\mathscr{C}(1)} \mathrm{d}(v)^{2-\theta}}{\big(\sum_{v\in\mathscr{C}(1)} \mathrm{d}(v)^{1-\theta}\big)^2} = O_{\mathbb{P}}^{\log N}\left(\frac{N^{(2-\theta)\gamma}}{N^{2\gamma}}\right) = O_{\mathbb{P}}^{\log N}(N^{-\theta\gamma}).$$

Hence, we obtain from (5.9) for $\theta > 0$

$$t^{\pi}_{\mathrm{meet}}(\mathscr{C}(1)) = \begin{cases} \Omega_{\mathbb{P}}^{\log N}(N^{\gamma}) & \text{if } \theta \ge 1, \\ \Omega_{\mathbb{P}}^{\log N}(N^{\gamma\theta}) & \text{if } \theta \in (0,1). \end{cases} \tag{5.10}$$

For the second of the part of the lower bound, we use a component that contains a sufficiently large "double star" structure and consider parameters $\theta < 1$. More precisely, by Proposition 3.2.7, with high probability, there exists a tree component that contains

two adjacent vertices x and y such that

$$\mathrm{d}(x), \mathrm{d}(y) \text{ and } \sum_{v \in \mathscr{C}(x)} \mathrm{d}(v) \text{ are } \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right). \tag{5.11}$$

Now, let A_x be the set of vertices in $\mathscr{C}(x)$ that are closer to x than to y, and A_y the complement. Then, we will use that by Proposition 4.1.4 (b)

$$\mathbb{E}_{\mu_u}(\tau_{\mathrm{cons}} \,|\, G_N) = \Omega\left(\frac{\pi(A_x)\pi(A_y)}{\sum_{v\in A_x}\sum_{w\in A_y} c(vw)}\right). \tag{5.12}$$

We start by estimating the term $\pi(A_x)\pi(A_y)$. Note that for $\theta \in (0,1)$, we have that

$$\mathrm{d}(x) \le |A_x| \le \sum_{v\in A_x} \mathrm{d}(v)^{1-\theta} \le \sum_{v\in\mathscr{C}(x)} \mathrm{d}(v)^{1-\theta} \le \sum_{v\in\mathscr{C}(x)} \mathrm{d}(v),$$

and the same bounds hold when replacing x by y. Therefore, for $\theta \in (0,1)$, we obtain

$$\frac{\mathrm{d}(x)}{\sum_{v\in\mathscr{C}(x)}\mathrm{d}(v)} \le \frac{\sum_{v\in A_x}\mathrm{d}(v)^{1-\theta}}{\sum_{v\in A_y}\mathrm{d}(v)^{1-\theta}} \le \frac{\sum_{v\in\mathscr{C}(x)}\mathrm{d}(v)}{\mathrm{d}(y)},$$

so that we can deduce from (5.11) that

$$\pi(A_x)\pi(A_y) = \left(\sqrt{\frac{\sum_{v\in A_x}\mathrm{d}(v)^{1-\theta}}{\sum_{v\in A_y}\mathrm{d}(v)^{1-\theta}}} + \sqrt{\frac{\sum_{v\in A_y}\mathrm{d}(v)^{1-\theta}}{\sum_{v\in A_x}\mathrm{d}(v)^{1-\theta}}}\right)^{-2} = \Omega_{\mathbb{P}}^{\log N}(1).$$

Furthermore, since $\mathscr{C}(x)$ is a tree the denominator in (5.12) reduces to

$$c(xy) = \frac{1}{\sum_{v\in\mathscr{C}(x)}\mathrm{d}(v)^{1-\theta}} \le \frac{1}{|\mathscr{C}(x)|} = O_{\mathbb{P}}^{\log N}\left(N^{-\frac{\gamma}{2-2\gamma}}\right).$$

We finally consider the case $\theta \le 0$ on this double star. Again, we start by estimating $\pi(A_x)\pi(A_y)$. By Proposition 3.2.7 we have

$$\sum_{v\in A_x} \mathrm{d}(v)^{1-\theta} = O_{\mathbb{P}}\Big(N^{\frac{\gamma(1-\theta)}{2-2\gamma}}\Big).$$

Since further $\mathrm{d}(x) = \Theta_{\mathbb{P}}(N^{\frac{\gamma}{2-2\gamma}})$ by (5.11), we also have that

$$\sum_{v\in A_x} \mathrm{d}(v)^{1-\theta} \ge \mathrm{d}(x)^{1-\theta} = \Omega_{\mathbb{P}}\Big(N^{\frac{\gamma(1-\theta)}{2-2\gamma}}\Big).$$

The same bounds hold for y and so by the same argument as above, we have that $\pi(A_x)\pi(A_y) = \Omega_{\mathbb{P}}(1)$. Moreover,

$$c(xy) = \frac{1}{\sum_{v\in\mathscr{C}(x)} \mathrm{d}(v)^{1-\theta}} \leq \frac{1}{\mathrm{d}(x)^{1-\theta}} = O_{\mathbb{P}}^{\log N}\Big(N^{\frac{-\gamma(1-\theta)}{2-2\gamma}}\Big).$$

Combining the estimates on the stationary distribution and the conductance $c(xy)$, we conclude from (5.12) that

$$\mathbb{E}_{\mu_u}(\tau_{\mathrm{cons}} \,|\, G_N) = \begin{cases} \Omega_{\mathbb{P}}^{\log N}(N^{\frac{\gamma}{2-2\gamma}}) & \text{if } \theta \in (0,1), \\ \Omega_{\mathbb{P}}^{\log N}(N^{\frac{\gamma(1-\theta)}{2-2\gamma}}) & \text{if } \theta \in (-\infty, 1]. \end{cases} \tag{5.13}$$

Combining Equations (5.13) and (5.10) completes the proof of Theorem 2.1.5 by giving all the required lower bounds. □

5.2 Consensus Time for the Discursive Voter Model

In this section, we will consider the discursive voter model as defined in Definition 2.1.3. This version of the voter model fits into the general setting of a Q-voter model of Section 4.2 if for $\theta \in \mathbb{R}$ we consider $Q = Q^\theta$ defined as

$$Q^\theta(i,j) = \frac{\mathrm{d}(i)^{\theta-1} + \mathrm{d}(j)^{\theta-1}}{2} \quad \text{if } i \sim j \text{ in } G_N. \tag{5.14}$$

As before, we write $\mathbb{P}^\theta$ for the law of (and $\mathbb{E}^\theta$ for the expectation with respect to) the coalescing random walks with generator Q^θ.

If we denote by $\mathscr{C}_1, \ldots, \mathscr{C}_k$ the connected components of G_N, then define $\pi = (\pi(z), z \in V(G_N))$ via

$$\pi(z) = \frac{1}{|\mathscr{C}_j|}, \quad \text{for } z \in \mathscr{C}_j,$$

for $j \in [k]$. Then $\pi|_{\mathscr{C}_j}$, i.e. the uniform measure on $\mathscr{C}_j$, is the invariant measure of the Q^θ Markov chain restricted to $\mathscr{C}_j$.

First we require another application of Theorem 4.1.10, which is simpler than for the classical voter model, but covers a wider range of cases.

Lemma 5.2.1. *For G_N with $\beta + 2\gamma < 1$, we have that*

$$t_{\text{meet}}(G_N) = \sup_{j\in[k]} t_{\text{meet}}(C_i) = \begin{cases} O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right) & \theta > \frac{3-4\gamma}{2-2\gamma} \\ O_{\mathbb{P}}^{\log N}\left(N^{\gamma(2-\theta)}\right) & 1 < \theta \leq \frac{3-4\gamma}{2-2\gamma} \\ O_{\mathbb{P}}^{\log N}\left(N^{\gamma}\right) & 2\gamma \leq \theta \leq 1 \\ O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(2-\theta)}{2-2\gamma}}\right) & \theta < 2\gamma \end{cases}$$

Proof. By Proposition 3.2.5, we can work on the high probability set where all big components are trees. Recall that $K_\gamma := N^{\frac{1-2\gamma}{2-2\gamma}} \log N$ and denote for any $k \leq K_\gamma$ by L_k the set of degree 1 vertices adjacent to k. By Proposition 3.2.3 (b) and since the stationary distribution is uniform

$$\min_{k\leq K_\gamma} \pi(L_k) = \min_{k\leq K_\gamma} \frac{|L_k|}{|\mathscr{C}(k)|} = \Omega_{\mathbb{P}}^{\log N}(1).$$

Then by exchangeability, coalescence observed in L_k is just complete graph (Wright-Fisher) coalescence. This is because a simultaneous move by both walkers gives the same probability to coalesce as a single move. Thus, coalescence occurs for the partially observed process at rate

$$\frac{1 + \mathrm{d}(k)^{\theta-1}}{|L_k|}$$

and we conclude by Proposition 3.2.3 (b)

$$\max_{k\leq K_\gamma} t^{\pi}_{\text{meet}}(L_k) \leq \max_{k\leq K_\gamma} \frac{|L_k|}{1 + \mathrm{d}(k)^{\theta-1}} = \begin{cases} O_{\mathbb{P}}^{\log N}(N^{\gamma(2-\theta)}) & \theta > 1, \\ O_{\mathbb{P}}^{\log N}(N^{\gamma}) & \theta \leq 1. \end{cases} \tag{5.15}$$

Now, we let $\mathfrak{S}$ be the collection of small components and branches in large components. If we denote by $P_{x,y}$ the set of paths between any vertices x and y, then by

Proposition 4.1.1 we obtain

$$\begin{aligned}\max_{S\in\mathfrak{S}} t_{\mathrm{hit}}(S) &\le \max_{S\in\mathfrak{S}} \max_{x,y\in V(S)} \min_{P_{x,y}} \sum_{\{u,v\}\in E(P_{x,y})} \frac{2|S|}{\mathrm{d}(u)^{\theta-1}+\mathrm{d}(v)^{\theta-1}} \\ &\le \max_{S\in\mathfrak{S}} |S|\operatorname{diam}(S) \max_{v\in S}\left(\mathrm{d}(v)^{1-\theta}\right) \\ &\le \max_{S\in\mathfrak{S}} |S|\operatorname{diam}(G_{\beta,\gamma}) \max_{v>K_\gamma}\left(\mathrm{d}(v)^{1-\theta}\right) \qquad (5.16)\\ &= \begin{cases} O_{\mathbb{P}}^{\log N}\left(N^{\frac{(2-\theta)\gamma}{2-2\gamma}}\right) & \theta<1, \\ O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right) & \theta\ge 1, \end{cases}\end{aligned}$$

where we used Lemma 3.2.5 and Propositions 3.2.2 (b) and 3.2.1 in the last step.

If we combine this last bound with (5.15) and apply Theorem 4.1.10, then we obtain

$$\begin{aligned}\max_{k\le K_\gamma} t_{\mathrm{meet}}(\mathscr{C}(k)) &= \begin{cases} O_{\mathbb{P}}^{\log N}\left(N^{\gamma(2-\theta)}+N^{\frac{\gamma}{2-2\gamma}}\right) & \theta>1, \\ O_{\mathbb{P}}^{\log N}\left(N^{\gamma}+N^{\frac{\gamma(2-\theta)}{2-2\gamma}}\right) & \theta\le 1, \end{cases} \\ &= \begin{cases} O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right) & \theta>\frac{3-4\gamma}{2-2\gamma}, \\ O_{\mathbb{P}}^{\log N}\left(N^{\gamma(2-\theta)}\right) & 1<\theta\le\frac{3-4\gamma}{2-2\gamma}, \\ O_{\mathbb{P}}^{\log N}\left(N^{\gamma}\right) & 2\gamma\le\theta\le 1, \\ O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(2-\theta)}{2-2\gamma}}\right) & \theta<2\gamma, \end{cases}\end{aligned}$$

which completes the proof of the lemma. □

Proof of Theorem 2.1.7. The upper bound for all four cases follows immediately from Lemma 4.2.4, Propositon 4.2.5 and Lemma 5.2.1, so it only remains to prove the lower bounds. For these, it will be very useful that the stationary distribution π on each component is always uniform.

The first lower bound is for the case $\theta \ge \frac{3-4\gamma}{2-2\gamma}$ for which we must consider the *long* double star component, whose existence is proved Proposition 3.2.8. First, we note that the 'separating edge' $\{v_2, v_3\}$ on the the long double star has conductance

$$c(v_2v_3) = O_{\mathbb{P}}^{\log N}\left(N^{-\frac{\gamma}{2-2\gamma}}\right).$$

Moreover, by Propositions 3.2.2 (a) and 3.2.3 (a)

$$\mathrm{d}(v_1), \mathrm{d}(v_4) \text{ and } |\mathscr{C}(v_1)| \text{ are } \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$$

which implies that we have $\Theta_{\mathbb{P}}^{\log N}(1)$ stationary mass on each side (by a similar argument as before). Hence by Proposition 4.1.4 (b) we have consensus time $\Omega_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right)$.

For the lower bound when $2\gamma < \theta < \frac{3-4\gamma}{2-2\gamma}$, we apply Corollary 4.1.4 (a) to $\mathscr{C}(1)$ to see that

$$t_{\mathrm{meet}}^{\pi}(\mathscr{C}(1)) \geq \frac{(1-\sum_{v\in\mathscr{C}(1)}\pi(v)^2)^2}{4\sum_{v\in\mathscr{C}(1)}q(v)\pi(v)^2} = \Theta_{\mathbb{P}}\left(\frac{|\mathscr{C}(1)|^2}{\sum_{v\in\mathscr{C}(1)}q(v)}\right) = \Theta_{\mathbb{P}}\left(\frac{|\mathscr{C}(1)|^2}{\sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^\theta}\right).$$

Recall the moment calculation in Lemma 3.2.6 to see that when $\theta \geq 1$

$$\sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^\theta = \Theta_{\mathbb{P}}^{\log N}\left(N^{\gamma\theta}\right),$$

whereas for $\theta \in (2\gamma, 1)$ we instead have by Proposition 3.2.2 (a)

$$\sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^\theta \leq \sum_{v\in\mathscr{C}(1)}\mathrm{d}(v) = O_{\mathbb{P}}^{\log N}(N^\gamma).$$

Combining the statements yields

$$\frac{|\mathscr{C}(1)|^2}{\sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^\theta} = \Omega_{\mathbb{P}}^{\log N}\left(N^{(2-\theta)\gamma} \vee N^\gamma\right). \tag{5.17}$$

By Lemma 4.2.4 and Proposition 4.2.5 this expression gives a lower bound for the consensus time.

For the final case, when $\theta < 2\gamma$, we require another double star component, but this one must be that without a path, whose existence is stated in Proposition 3.2.7. This double star is a tree structure with two adjacent "star" vertices x, y, where

$$\mathrm{d}(x), \mathrm{d}(y) \text{ and } |\mathscr{C}(x)| \text{ are } \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right).$$

Therefore we have stationary mass of $\Theta_{\mathbb{P}}^{\log N}(1)$ in the vertices closest to x and in those closest to y. We note that

$$Q^\theta(x,y) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(\theta-1)}{2-2\gamma}}\right)$$

and since the stationary distribution is uniform

$$\pi(x) = O_{\mathbb{P}}^{\log N}\left(N^{-\frac{\gamma}{2-2\gamma}}\right).$$

Thus, by the definition of the conductance (4.1)

$$c(xy) = \pi(x)Q^{\theta}(x,y) = O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(\theta-1)}{2-2\gamma}}N^{-\frac{\gamma}{2-2\gamma}}\right).$$

Combining the estimate on the stationary mass and the conductance, we have by Proposition 4.1.4 (b),

$$t_{\text{meet}}(\mathscr{C}(1)) = \Omega_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma(2-\theta)}{2-2\gamma}}\right),$$

which gives the last remaining lower bound. □

5.3 Extending to the Erdős-Rényi Case

We would like to see the main consensus theorems as valid for $\gamma \in [0,1)$. So, in this section we extend the previous proofs to the $\gamma = 0$ case, the Erdős-Rényi network $\mathscr{G}(n,p)$ with homogenous edge probabilities

$$p_{ij} = \frac{\beta}{N}$$

for the subcritical parameter $\beta < 1$. We use this usual definition rather than the $1 - e^{-\beta/N}$ of Definition 2.0.4 because it's simpler, and the two definitions are asymptotically equivalent (they are both in the class of Definition 2.0.5).

Proposition 5.3.1. *For G_N with $\beta > 0$, $\gamma = 0$ and any $C > 2$ we have*

$$\max_{v\in[N]} \mathrm{d}(v) \leq C\log N$$

with high probability as $N \to \infty$.

Proof. Any $v \in [N]$ has

$$\mathrm{d}(v) \overset{\mathrm{d}}{=} \mathrm{Bin}\left(N-1, \frac{\beta}{N}\right) \preceq \mathrm{Bin}\left(N, \frac{\beta}{N}\right)$$

so we can use a standard Chernoff bound with $p > 0$ to obtain

$$\mathbb{P}\left(\frac{1}{N}\,\mathrm{d}(v) \geq \frac{\beta}{N} + p\right) \leq \exp\left(-\frac{Np^2}{2\beta/N + 2p}\right).$$

Therefore if we take $p = C\frac{\log N}{N}$ we find this probability is $o\left(\frac{1}{N}\right)$ whenever $C > 2$ and so we can claim the bound jointly for every $v \in [N]$ with high probability, by the union bound. Finally we move to the claim of the proposition by noting that we can take any $C' \in (2, C)$ and find

$$\mathbb{P}\left(\frac{1}{N}\,\mathrm{d}(v) \geq C\frac{\log N}{N}\right) \leq \mathbb{P}\left(\frac{1}{N}\,\mathrm{d}(v) \geq \frac{\beta}{N} + C'\frac{\log N}{N}\right) = o\left(\frac{1}{N}\right)$$

where the first inequality holds by taking N sufficiently large. □

We also bound the component size with the following proposition which follows immediately from [Bollobás et al., 2007, Theorem 3.12(i)].

Proposition 5.3.2. *For G_N with $\beta \in (0,1)$ and $\gamma = 0$ we find*

$$|\mathscr{C}_{\max}| = O_{\mathbb{P}}(\log N)$$

Corollary 5.3.3. *For Discursive or Classical dynamic dual walkers with any parameter $\theta \in \mathbb{R}$ on G_N with $\gamma = 0$ and $\beta \in (0,1)$ we can lower bound the minimum ergodic flow*

$$\min_{i \sim j} c(i,j) = \Omega_{\mathbb{P}}^{\log N}(1)$$

where as usual $c(i,j) := \pi(i)Q(i,j)$ refers to the stationary distribution for the chain restricted to $\mathscr{C}(i)$.

Proof. Write $D := \max_{v \in [N]} \mathrm{d}(v)$, $C := |\mathscr{C}_{\max}|$ – both quantities are $O_{\mathbb{P}}^{\log N}(1)$ by the previous propositions. For the Classical dynamic when $i \sim j$

$$c(i,j) = \pi(i)\,\mathrm{d}(i)^{\theta-1} = \frac{1}{\sum_{v \in \mathscr{C}(i)} \mathrm{d}(v)^{1-\theta}}$$

which we can loosely bound at

$$\sum_{v \in \mathscr{C}(i)} \mathrm{d}(v)^{1-\theta} \leq C \cdot \left(1 \wedge D^{1-\theta}\right) = O_{\mathbb{P}}^{\log N}(1).$$

Instead for the Discursive dynamic when $i \sim j$

$$c(i,j) = \pi(i)\left(\frac{\mathrm{d}(i)^{\theta-1} + \mathrm{d}(j)^{\theta-1}}{2}\right) \geq \frac{1}{C} \cdot \left(1 \vee D^{\theta-1}\right) = \Omega_{\mathbb{P}}^{\log N}(1)$$

by recalling that the stationary distribution is uniform on components for this model. □

Theorem 5.3.4. *Take $\beta \in (0,1)$ and $\gamma = 0$ with initial conditions distributed as μ_u such that each initial opinion is an independent* Bernoulli(u) *random variable, for some $u \in (0,1)$. Then, for the classical voter model on G_N with parameter $\theta \in \mathbb{R}$, we have*

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\mathrm{cons}} | G_N) = O_{\mathbb{P}}^{\log N}(1)\,. \tag{5.18}$$

Proof. Proposition 5.3.2 trivially bounds the componentwise diameter $\mathrm{diam}\,(G_N) = O_{\mathbb{P}}(\log N)$ and Corollary 5.3.3 bounds the conductance. Hence by Proposition 4.1.1 we find

$$t_{\mathrm{hit}} = O_{\mathbb{P}}^{\log N}(1)$$

which gives the same upper bound on the expected consensus time via Proposition 4.2.5. □

5.4 Comparison to the Largest Component

Returning to the scale-free context, we have two complementary results for the consensus time on the largest component (see Definition 2.1.2) in Propositions 2.1.6 and 2.1.8. These results do not require substantially new ideas to prove, but still their proof is not contained in the proofs of the main theorems so we will prove them instead in this section.

Proposition 5.4.1. *When $\gamma \in \left(0, \frac{1}{2}\right)$ and $\beta > 0$ have $\beta + 2\gamma < 1$, the largest branch (see Definition 3.2.4) of $\mathscr{C}(1)$ has $O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma^2}{1-\gamma}}\right)$ vertices.*

Proof. $\mathscr{C}(1)$ is a tree with high probability by Lemma 3.2.5, and as in Section 3.2 we will contain this tree in the unthinned Galton-Watson process. By Proposition 3.2.3 (a)

$$\mathrm{d}(1) = \Theta_{\mathbb{P}}\left(N^{\gamma}\right)$$

from which point we have $\Theta_{\mathbb{P}}\left(N^{\gamma}\right)$ iid subcritical Galton-Watson trees $(T_i)_{i=1}^{\mathrm{d}(1)}$. We can

bound the tail of the distribution of these trees using Proposition 3.2.10

$$\mathbb{P}\left(|T_i| > N^{\frac{\gamma^2}{1-\gamma}} \log N\right) = O\left(\left(N^{\frac{\gamma^2}{1-\gamma}} \log N\right)^{1-\frac{1}{\gamma}}\right) = o\left(N^{-\gamma}\right)$$

and so by the union bound conditionally on $\mathrm{d}(1) = \Theta_{\mathbb{P}}\left(N^{\gamma}\right)$, with high probability none of these branches will exceed the limit $N^{\frac{\gamma^2}{1-\gamma}} \log N$. □

We can provide the corresponding lower bound on the branch size by finding an appropriate neighbour of 1.

Proposition 5.4.2. *When $\gamma \in \left(0, \frac{1}{2}\right)$ and $\beta > 0$ have $\beta + 2\gamma < 1$, with high probability there is some $i \in [N]$ such that $i \sim 1$ in G_N and $\mathrm{d}(i) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma^2}{1-\gamma}}\right)$. Further, i has $\Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma^2}{1-\gamma}}\right)$ neighbours of degree 1.*

Proof. Define $L := \left\lfloor N^{\frac{1-2\gamma}{1-\gamma}} \log N \right\rfloor$. The MNR Poisson weight between 1 and $[L] \setminus \{1\}$ is

$$\sum_{j=2}^{\lfloor L \rfloor} \beta N^{2\gamma-1} j^{-\gamma} \sim \frac{\beta}{1-\gamma} N^{2\gamma-1} L^{1-\gamma} \to \infty$$

as $N \to \infty$. Hence, because a $\mathrm{Pois}(\mu)$ distribution puts $e^{-\mu}$ mass at 0, the probability that there is no such edge tends to 0 with N. We further calculate as in Equation (3.16) that $\mathscr{C}(1)$ expects no more shared indices than

$$\mathbb{P}(M = M') N^{2\gamma} \log^2 N = \Theta_{\mathbb{P}}^{\log N}\left(N^{2\gamma-1}\right) \to 0$$

and so with high probability we will not have to thin the Galton-Watson exploration at all in constructing $\mathscr{C}(1)$. Hence, the degree of this neighbour will on this high probability assumption have

$$\mathrm{d}(i) \stackrel{\mathrm{d}}{=} 1 + \mathrm{Pois}(w(i)) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma^2}{1-\gamma}}\right)$$

and in the next generation of the Galton-Watson exploration we will have a Binomial number of individuals with no offspring, which proves the final claim. □

Proof of Proposition 2.1.6. First we recall the claimed exponents

$$\mathbb{E}^{\theta}_{\mu_u}(\tau_{\text{cons}}(\mathscr{C}(1))|G_N) = \Theta^{\log N}_{\mathbb{P}}\left(N^c\right), \quad \text{where } c = \begin{cases} \gamma & \theta \geq 1, \\ \gamma\theta & \frac{\gamma}{1-\gamma} < \theta < 1, \\ \frac{\gamma^2}{1-\gamma} & 0 \leq \theta \leq \frac{\gamma}{1-\gamma}, \\ \frac{\gamma^2(1-\theta)}{1-\gamma} & \theta < 0. \end{cases}$$

The lower bounds for the first two cases are already stated in Equation (5.10) as they also provide lower bounds for the whole network's consensus time.

For the second two cases, the lower bounds are from slow mixing with the conductance result of Proposition 4.1.4 (b). Consider the vertex i which we found in Proposition 5.4.2. If we cut the component at the edge $\{1, i\}$ and write A_1, A_i for the two halves then we find

$$\mathbb{E}_{\mu_u}(\tau_{\text{cons}}\left(\mathscr{C}(1)\right) \mid G_N) = \Omega_{\mathbb{P}}\left(\frac{\pi(A_i)\pi(A_1)}{c(1i)}\right).$$

Because $\pi(A_1) = \Theta_{\mathbb{P}}(1)$ we can discard this factor. Now

$$\frac{\pi(A_i)}{c(1i)} = \frac{\sum_{v\in A_i}\mathrm{d}(v)^{1-\theta}}{\sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^{1-\theta}} \cdot \sum_{v\in\mathscr{C}(1)}\mathrm{d}(v)^{1-\theta} = \sum_{v\in A_i}\mathrm{d}(v)^{1-\theta} = \Omega_{\mathbb{P}}\left(N^{\frac{\gamma^2}{1-\gamma}} + N^{\frac{\gamma^2(1-\theta)}{1-\gamma}}\right)$$

where the final claim is from considering the sum of the degree at i and of its degree 1 neighbours mentioned in Proposition 5.4.2. This completes the lower bounds.

In the first case $\theta \geq 1$, the upper bound can simply be from the G_N consensus time of Theorem 2.1.5 which is also $\Theta^{\log N}_{\mathbb{P}}\left(N^{\gamma}\right)$.

For the middle two cases, the upper bounds follow from Equation (5.7) together with inserting our new branch exponent (from Proposition 5.4.1) into the calculation in (5.8). This obtains

$$t_{\text{meet}}(\mathscr{C}(1)) = O^{\log N}_{\mathbb{P}}\left(N^{\gamma\theta} + N^{\frac{\gamma^2}{1-\gamma}}\right)$$

for these parameters $\theta \in (0, 1)$.

For the final upper bound, in the case $\theta \leq 0$, we use Proposition 4.1.7 to bound the meeting time by $t_{\text{hit}}(1)$. Take some $v \in \mathscr{C}(1)$ in some branch B. Then we calculate in the same way as in (5.3) to see

$$\mathbb{E}_v\left(T_1\right) \leq \text{diam}(B)\sum_{w\in B}\mathrm{d}(w)^{1-\theta} \leq \text{diam}(B)\left(\sum_{w\in B}\mathrm{d}(w)\right)^{1-\theta} = O_{\mathbb{P}}\left(N^{\frac{\gamma^2(1-\theta)}{1-\gamma}}\right)$$

which concludes the upper bounds. □

To prove the Lemma for $\mathscr{C}(1)$ consensus times with the discursive dynamic, we must first describe a different branch of $\mathscr{C}(1)$ which also exists with high probability in the structure.

Proposition 5.4.3. *When $\gamma \in \left(0, \frac{1}{2}\right)$ and $\beta > 0$ have $\beta + 2\gamma < 1$, with high probability there is some $j \in [N]$ such that $\mathrm{d}(1, j) = 3$ in G_N, $\mathrm{d}(j) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma^2}{1-\gamma}}\right)$ and the unique path connecting 1 to j features just two vertices of degree 2.*

Proof. Construct $\mathscr{C}(1)$ using the Galton-Watson tree as in Proposition 5.4.2, and recall again that this tree will with high probability not require any thinning. Our offspring distribution from Section 3.1 has probability $p_1 > 0$ of having a single offspring because it is mixed Poisson, and so if we explore the ball of radius 2 around 1 we will find the number of branches which appear as a simple path of length 3 is concentrated around

$$p_1^2 \, \mathrm{d}(1) = \Theta_{\mathbb{P}}^{\log N}\left(N^{\gamma}\right).$$

Continuing this exploration to the third generation of offspring, at least one of these edges will find $[L] \setminus \{1\}$ with $L := \left\lfloor N^{\frac{1-2\gamma}{1-\gamma}} \log N \right\rfloor$ and we conclude just as in the previous construction of Proposition 5.4.2. □

Proof of Proposition 2.1.8 . First we recall the claimed exponents

$$\mathbb{E}_{\mu_u}^{\theta}(\tau_{\mathrm{cons}}(\mathscr{C}(1)) | G_N) = \Theta_{\mathbb{P}}^{\log N}\left(N^{c}\right), \quad \text{where } c = \begin{cases} \frac{\gamma^2}{1-\gamma} & \theta \geq \frac{2-3\gamma}{1-\gamma}, \\ \gamma(2-\theta) & 1 < \theta < \frac{2-3\gamma}{1-\gamma}, \\ \gamma & 3 - \frac{1}{\gamma} \leq \theta \leq 1, \\ \frac{\gamma^2(2-\theta)}{1-\gamma} & \theta < 3 - \frac{1}{\gamma}. \end{cases}$$

Here we already proved a lower bound for the *middle* two cases which we see in Equation (5.17). For the final case $\theta < 3 - \frac{1}{\gamma}$ we consider the large vertex $i \sim 1$ of Proposition 5.4.2 just as in the proof for the classical dynamic, cutting the component at the edge $\{1, i\}$ and writing A_1, A_i for the two halves. Then by Proposition 4.1.4 (b)

$$\mathbb{E}_{\mu_u}(\tau_{\mathrm{cons}}\left(\mathscr{C}(1)\right) \mid G_N) = \Omega_{\mathbb{P}}\left(\frac{\pi(A_i)\pi(A_1)}{c(1i)}\right).$$

and the uniform stationary distribution makes calculation easy

$$\pi(A_i)\cdot\pi(A_1)\cdot\frac{1}{c(1i)}=\frac{|A_i|}{|\mathscr{C}(1)|}\cdot\frac{|A_1|}{|\mathscr{C}(1)|}\cdot\frac{|\mathscr{C}(1)|}{Q(1,i)}=\frac{2|A_i|}{d(i)^{\theta-1}}\cdot\Theta_{\mathbb{P}}(1)=\Theta_{\mathbb{P}}\left(N^{\frac{\gamma^2(2-\theta)}{1-\gamma}}\right).$$

We've used here that $\theta-1<2-\frac{1}{\gamma}<0$ to say

$$Q(1,i)=\frac{d(1)^{\theta-1}+d(i)^{\theta-1}}{2}\overset{\mathbb{P}}{\sim}\frac{d(i)^{\theta-1}}{2}.$$

We still have to prove the first lower bound in the case $\theta\geq\frac{2-3\gamma}{1-\gamma}$, and for that we require the new branch of $\mathscr{C}(1)$ given by Proposition 5.4.3. Here, let the path from 1 to j be labelled

$$1\leftrightarrow u\leftrightarrow v\leftrightarrow j$$

and remove the edge u,v between 2 vertices of degree 2. Then

$$\mathbb{E}_{\mu_u}(\tau_{\mathrm{cons}}(\mathscr{C}(1))\mid G_N)=\Omega_{\mathbb{P}}\left(\frac{\pi(A_u)\pi(A_v)}{c(uv)}\right)$$

and we find in this expression the final lower bound

$$\frac{\pi(A_u)\pi(A_v)}{c(uv)}=\frac{|A_u||A_v|}{|\mathscr{C}(1)|^2}\frac{|\mathscr{C}(1)|}{2^{\theta-1}}=\Theta_{\mathbb{P}}\left(N^{\frac{\gamma^2}{1-\gamma}}\right).$$

We now consider the upper bounds using the partial meeting result Theorem 4.1.10 by looking at the partially observed chain on the set L_1 of leaves adjacent to 1. The upper bound for $t^{\pi}_{\mathrm{meet}}(L_1)$ is already done in Equation (5.15) so it remains to bound $t_{\mathrm{hit}}(1)$, for which we can repeat Equation (5.16) but only looking at the branches of $\mathscr{C}(1)$. Recalling the maximal size of these branches from Proposition 5.4.1, we find

$$t_{\mathrm{hit}}(1)=\begin{cases}O_{\mathbb{P}}^{\log N}\left(N^{\frac{(2-\theta)\gamma}{2-2\gamma}}\right) & \theta<1,\\ O_{\mathbb{P}}^{\log N}\left(N^{\frac{\gamma}{2-2\gamma}}\right) & \theta\geq 1,\end{cases}$$

which combines with Equation (5.15) in Theorem 4.1.10 to obtain

$$t_{\mathrm{meet}}(\mathscr{C}(1))=\begin{cases}O_{\mathbb{P}}^{\log N}\left(N^{\gamma}+N^{\frac{(2-\theta)\gamma}{2-2\gamma}}\right) & \theta<1,\\ O_{\mathbb{P}}^{\log N}\left(N^{(2-\theta)\gamma}+N^{\frac{\gamma}{2-2\gamma}}\right) & \theta\geq 1,\end{cases}$$

completing the proof. □

Chapter 6

Voter Model Mixing on a Supercritical Scale-Free Network

In the previous chapter we carried out a complete analysis of these voter models in the case of the subcritical network, and so it is natural to next ask what we can say about the supercritical case. Hence in this chapter we consider the regime $\beta + 2\gamma > 1$ where the giant component exists, with our attention restricted to the discursive family of voter models which turn out to be more tractable in this case. Primarily, they are more tractable because the symmetry of the generator means that relaxation times are monotonically decreasing with θ and so bounding the mixing time of one voter model has consequences for all larger θ parameters. Note that now, as $\gamma \geq \frac{1}{2}$ parameters are included in this set, the choice of SNR network definition becomes relevant as for those parameters it is no longer equivalent to the CL definition.

The discursive voter model has a uniform stationary distribution on the giant component for any parameter $\theta \in \mathbb{R}$, and so by applying the hitting time bound using Lemma 3.3.5 (which gives paths through low degree vertices) we quickly see that consensus is $O_{\mathbb{P}}^{\log N}(N)$ for any θ. We can only match this with a lower bound, and conclude that consensus time is $\Theta_{\mathbb{P}}^{\log N}(N)$, when $\theta\gamma \leq 1$. The proof for consensus time order in this case is at the end of this chapter.

In investigating the parameters $\theta > \frac{1}{\gamma}$ we managed to prove a polylogarithmic mixing time bound that applies to the dual chain of discursive models with $\theta \geq 1$, on the condition that β is sufficiently large. However, results [Chen et al., 2016, (3.21)] and [Aldous and Fill, 2002, Lemma 3.17], with discrete-time analogues in [Kanade et al.,

2016], only give the consensus bounds

$$\Omega_{\mathbb{P}}^{\log N}\left(\frac{1}{\sum_v q(v)\pi(v)^2}\right), \quad O_{\mathbb{P}}^{\log N}\left(\frac{t_{\text{mix}}}{\sum_v \pi(v)^2}\right),$$

which are not polylogarithmically tight in this second case $\theta\gamma > 1$.

Instead from this mixing result we derive Corollary 2.1.13 by duality, which tells us that the voter model has at the slowest particular vertex a time correlation decaying on a polylogarithmic time frame. Therefore in this voter model problem, initial opinions are forgotten very quickly relative to the total lifetime of the process and replaced instead with something closer to a uniform opinion drawn from the whole initial condition.

Transitions in the edge density β, excluding those caused by the appearance of the giant component, are rare for interacting particle systems on networks and so the large β condition should not be thought of as the important omission. Regardless, while the proof could be adapted to allow for smaller β, the techniques require the artificial condition $\beta > 1$. This is because the mixing time is bounded by finding an Erdős-Rényi giant component as a subgraph and growing it to a subgraph spanning the giant component of G_N.

This approach is inspired by the structural theorems in [Fountoulakis and Reed, 2008] and [Benjamini et al., 2014] which led to mixing time bounds for the Erdős-Rényi giant component. It is also similar in some ways to the approach of Theorem 3.3.4 in using the wealth of research on the Erdős-Rényi giant component applied to the subgraph of this type in the SNR model for which less research has been done.

A monotonicity in θ of these models will allow all mixing results to follow from a bound on the mixing time of the variable speed random walk on the SNR graph, and hence most of the chapter will be dedicated to proving that bound. Of course the variable speed random walk is a natural model and so this bound has some independent interest.

Future Work

There are many plausible heuristic arguments that could apply to the $\theta\gamma > 1$ case. The interacting system on the giant component is not invariant to permutations of the agents and so we cannot properly take the mean-field limit as in [Mach et al., 2020]. However, due to the uniform stationary distribution, permutations do not affect the total opinion martingale (M_t) of Proposition 4.2.1 so we might still think taking the mean field approximation, by treating site opinions as independent, would give a good

prediction. This leads to

$$\frac{\mathrm{d}\left(M_t^2\right)}{\mathrm{d}t} = M_t(1-M_t)\sum_v \pi^2(v)q(v)$$

so we identify the Wright-Fisher diffusion of Definition 4.2.2 on the timescale of the lower bound of Proposition 4.1.4 (a). The order of this timescale is

$$\frac{1}{\sum_{v\in\mathscr{C}_{\max}}\pi(v)^2 q(v)} = \frac{|\mathscr{C}_{\max}|^2}{\sum_{v\in\mathscr{C}_{\max}}\mathrm{d}(v)^\theta} \sim \begin{cases} N & \theta \leq \frac{1}{\gamma} \\ N^{2-\theta\gamma} & \theta > \frac{1}{\gamma} \end{cases} \tag{6.1}$$

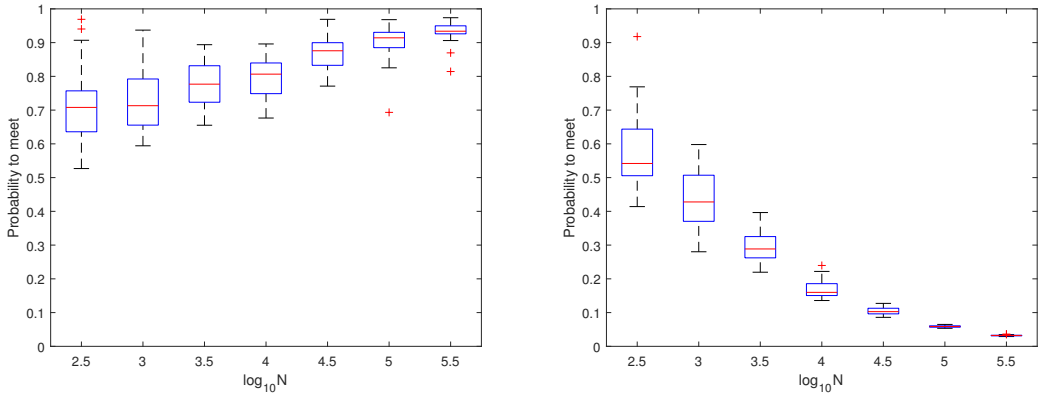
which would be our conjecture for the consensus time if we believe the mean-field approximation is working well. Of course, it cannot therefore be working well in this sense if $\theta\gamma > 2$.

In the classical dynamic when $\theta = 0$, [Durrett, 2010] conjectures via Aldous' "Poisson Clumping Heuristic" [Aldous, 1989] that the order of the mean consensus time is really the exact polynomial without logarithmic corrections as was also found non-rigorously in [Sood et al., 2008]. This heuristic uses Kac's formula [Aldous and Fill, 2002, (2.24)] for the return time to the "diagonal" set $\{(i,i) : i \in \mathscr{C}(1)\}$. Applied to any voter model with dual chain having stationary distribution π and vertex rates q, it tells us that for the distribution $\rho \propto q$ and some point $P \sim \rho$

$$\mathbb{E}_{(P,P)}\left(t_{\text{meet}}\right) = \frac{1-\sum_v \pi^2(v)}{2\sum_v \pi^2(v)q(v)}$$

which is the same expression as [Chen et al., 2016, (3.17)] and the Proposition 4.1.4 (a) order again. So, if we have an asymptotically positive probability to escape from (P,P) (in the sense of having a strong stationary time before returning to the diagonal), then we might argue that $\mathbb{E}_{(P,P)}\left(t_{\text{meet}}\right)$ has the same order as $\mathbb{E}_{\pi\otimes\pi}\left(t_{\text{meet}}\right)$, which is what [Chen et al., 2016, Theorem 2.2] identify as the voter model consensus time.

In the case $\theta\gamma > 1$, however, vertices of large degree have high rates in both directions on their incident edges and so it's quite plausible that the probability to meet in $O(1)$ time from (P,P) is $1 - o_{\mathbb{P}}^{\log N}(1)$. We simulate this meeting probability in 6.1 and it appears that in fact the answer might depend on more than just $\theta\gamma$: based on these simulations we conjecture that the mean-field approximation is best suited to the case $\gamma > \frac{1}{2}$ and so it would be only in $\{\theta\gamma \geq 1\} \cap \{\gamma \geq \frac{1}{2}\}$ that we see the exponent $2-\theta\gamma$. However, we've learnt from [Pastor-Satorras and Vespignani, 2002] that effects of large vertices could well appear later in the limit.



(a) $\gamma = \frac{1}{4}$ (b) $\gamma = \frac{3}{4}$

Figure 6.1: On the giant component of an SNR network with $\theta\gamma = \frac{3}{2}$, we show the probability of two particles initially at (P, P) meeting before they are separated by distance 3, as it depends on the graph size N. These box plots show just the random effect of the graph (for each realisation, meeting probability calculation is done to precision 0.01).

As well as settling the remainder of the discursive picture, in the future we might also like to settle the consensus time for the family of classical models. A very general rigorous analysis for voter model consensus time is made in [Chen et al., 2016], under assumption on the mixing and meeting times for the dual chain. However, mixing times in particular are very sensitive to work with and therefore highly model-dependent. In particular, they are open for this class $\mathscr{G}_{\beta,\gamma}$ of rank-one network models even for the SRW/CSRW and so the conjectures of [Durrett, 2010] are all we have.

Of course $\mathscr{G}_{\beta,\gamma}$ contains configuration models and on configuration models we have some mixing results. However, existing results assume *subpolynomial maximum degree* as in [Berestycki et al., 2018], or a *degree lower bound* as in [Abdullah et al., 2012]. The conjecture then is that these results do extend to general configuration models with power-law degree sequence and so $t_{\text{mix}} = \Theta_{\mathbb{P}}\left(\log^2 N\right)$. A structural result comparable to [Ding et al., 2014] but for the rank one scale-free network would solve the open question of mixing time, but also potentially give a direct handle on meeting time without the logarithmic factors.

6.1 Mixing Time for the Variable Speed Random Walk

We denote by the variable speed random walk (VSRW) the second most popular definition of the "simple random walk" on a graph. This is the Markov chain with

$$Q(i,j) = \mathbb{1}_{i\sim j}$$

for every pair $i \neq j$ in the vertex space. The VSRW is the dual Markov chain for both discursive and classical voter models when $\theta = 1$ – the two voting dynamics are identical at this parameter. The jump chain for the VSRW is the discrete time SRW. However, we cannot call it the "continuous time simple random walk" as this is reserved for the version with rates at each vertex i scaled by $\frac{1}{\mathrm{d}(i)}$, called the constant speed simple random walk (CSRW), which has equivalent mixing times to the discrete time SRW.

Despite the fact that the VSRW is a timechange of the CSRW which strictly accelerates, in the literature there does not seem to be a general mixing time relationship between the two and it's not immediately clear if such a relation should exist.

The voter model correlation result Theorem 2.1.13 proved in this chapter relies primarily on proving polylogarithmic mixing times for the VSRW on the simple Norros-Reittu network, claimed in Theorem 2.1.12 which we prove in this section.

As well as the relaxation and mixing times defined in Section 4.1.2, we introduce the Cheeger constant which is useful for understanding mixing of the simple random walk on a simple graph, as it controls over every subset the proportion of the ergodic flow not leading out of that subset.

Definition 6.1.1 (Cheeger constant)**.** For any connected graph G on $[N]$ we define the (VSRW) Cheeger constant

$$\Phi(G) := \inf_{\substack{S\subset[N]\\ |S|\leq N/2}} \frac{e\left(S : S^C\right)}{|S|}$$

where $e\left(A : B\right)$ denotes the number of edges incident to a vertex in both A and B, i.e. the number of edges between sets A and B.

The theory for Cheeger constants was originally written out in discrete time in [Sinclair, 1992]. In [Aldous and Fill, 2002] these are translated to continuous time so again we use that for reference.

Theorem 6.1.2. *The VSRW relaxation time for a connected graph G on $[N]$ has*

$$\frac{1}{2\Phi(G)} \leq t_{\text{rel}} \leq \frac{8}{\Phi(G)^2} \max_{v\in[N]} \mathrm{d}(v)$$

Proof. Cheeger's inequality [Aldous and Fill, 2002, Theorem 4.40] gives

$$t_{\text{rel}} \leq 8\tau_c^2 \max_{v\in[N]} \mathrm{d}(v)$$

for the different parameter τ_c which has

$$\tau_c := \sup_{A\subset[N]} \frac{|A|}{e(A:A^c)} \cdot \frac{N-|A|}{N} \leq \sup_{\substack{A\subset[N] \\ |A|\leq\frac{1}{2}}} \frac{|A|}{e(A:A^c)} = \frac{1}{\Phi(G)}.$$

For the other direction, this τ_c definition is symmetric so we can assume w.l.o.g. that $|A| \leq N - |A|$. Then [Aldous and Fill, 2002, Corollary 4.37] completes the proof:

$$t_{\text{rel}} \geq \tau_c \geq \frac{1}{2} \sup_{A\subset[N]} \frac{|A|}{e(A:A^c)} = \frac{1}{2\Phi(G)}.$$

□

Lemma 6.1.3. *If G is the giant component of an Erdős-Rényi graph on $[N]$ with parameter $\beta > 1$ (see Definition 2.2.3) then*

$$\Phi(G) = \Omega_{\mathbb{P}}\left(\frac{1}{\log^2 N}\right)$$

Proof. From [Benjamini et al., 2014] we have *for the CSRW*

$$t_{\text{mix}}(G) = O_{\mathbb{P}}\left(\log^2 N\right)$$

and Lemma 4.1.3 translates this to the same bound on $t_{\text{rel}}(G)$. Then [Aldous and Fill, 2002, Corollary 4.37] applies to τ_c for the CSRW which has form

$$\tau_c := \sup_{A\subset[N]} \frac{\pi(A)\pi(A^c)}{\sum_{x\in A, y\notin A} \pi(x)Q(x,y)} = \frac{1}{\mathrm{d}([N])} \sup_{A\subset[N]} \frac{\mathrm{d}(A)\,\mathrm{d}(A^c)}{e(A:A^c)}$$

and so we can deduce

$$t_{\text{rel}} \geq \tau_c \geq \frac{1}{2} \sup_{A\subset[N]} \frac{\mathrm{d}(A)}{e(A:A^c)} \geq \frac{1}{2} \sup_{A\subset[N]} \frac{|A|}{e(A:A^c)} = \frac{1}{2\Phi(G)}$$

which after inversion is the desired bound. □

We also require the following simple degree bound.

Lemma 6.1.4. *For the SNR network G_N with any parameters $\beta > 0$, $\gamma \in (0,1)$ we find, for any $\epsilon > 0$,*

$$\max_{v \geq \epsilon N \log^{-\frac{1}{\gamma}} N} \mathrm{d}(v) = O_{\mathbb{P}}(\log N).$$

Proof. Recall that the Norros-Reittu weights $w(v)$ satisfy

$$w(v) := \beta v^{-\gamma} \sum_{x=1}^{N} x^{-\gamma} < \beta v^{-\gamma} \int_0^N x^{-\gamma} \mathrm{d}x = \frac{\beta}{1-\gamma} \left(\frac{N}{v}\right)^{\gamma}$$

so that

$$\max_{v \geq \epsilon N \log^{-\frac{1}{\gamma}} N} w(v) < \frac{\beta \epsilon^{-\gamma}}{1-\gamma} \log N.$$

Recall also that in the MNR version we have exactly degrees with Poisson distribution according to their MNR weights. Hence, because the SNR version is constructed from the MNR version by flattening, we have the stochastic order

$$\mathrm{d}(v) \preceq \mathrm{Pois}\left(\frac{\beta \epsilon^{-\gamma}}{1-\gamma} \log N\right).$$

If we fix a constant C with $e^C - 1 = \frac{1-\gamma}{\beta \epsilon^{-\gamma}}$ then we conclude

$$\mathbb{E}\left(e^{C\,\mathrm{d}(v)}\right) < N,$$

and so by the Chernoff bound we deduce for any $a > 2$ that $C\,\mathrm{d}(v) < a \log N$ with probability $1 - o\left(\frac{1}{N}\right)$. The result follows by the union bound. □

We can now prove the main result of this section bounding the VSRW mixing time on an SNR network.

Proof of Theorem 2.1.12. Recall that the MNR network is constructed via the kernel

$$f : (x, y) \mapsto \beta x^{-\gamma} y^{-\gamma}$$

in that there are $\mathrm{Pois}\left(\frac{1}{N} f\left(\frac{v}{N}, \frac{w}{N}\right)\right)$ edges between each pair $v, w \in [N]$. Thinking of f as the density of a measure on $(0,1]^2$, we can decompose its mass into a sum of constituent parts which we will realise independently to construct a subgraph of the MNR network.

For some $\epsilon > 0$, we distinguish the vertex set of high weight vertices

$$H_\epsilon = \left[\left\lfloor \epsilon N \log^{-\frac{1}{\gamma}} N \right\rfloor\right]$$

and assume that $\beta > 1$ so that we can use the constant part $(x, y) \mapsto \beta$ to construct an Erdős-Rényi graph with probabilities β/N on the set $[N] \setminus H_\epsilon$. In fact, this part generates the graph with edge probabilities

$$p_{ij} = \begin{cases} 1 - e^{-\beta/N} & i, j \notin H_\epsilon \\ 0 & \text{otherwise,} \end{cases}$$

but by asymptotic equivalence there is whp no difference between this definition and the definition β/N [van der Hofstad, 2016, Theorem 6.18].

In this Erdős-Rényi graph, realise the largest component $\mathscr{G}_1$ on a vertex set denoted $\mathscr{C}_1$. By constructing the largest component, we have conditioned that the other components are smaller. We note further (see e.g. [Bollobás et al., 2007], using that $|H_\epsilon| = o(N)$) that

$$\frac{|\mathscr{C}_1|}{N} \xrightarrow{\mathbb{P}} \rho$$

where $\rho \in (0, 1)$ is the unique solution to $\rho + e^{-\beta\rho} = 1$. Hence this is the giant component, and we find for any small $\delta \in \left(0, \frac{1}{2}\right)$ that $|\mathscr{C}_1| \geq (1 - \delta)\rho N$ with high probability.

By using that

$$1 - \rho = e^{-\beta\rho} < \frac{1}{1 + \beta\rho} \implies \beta(1 - \rho) < 1,$$

we can take δ sufficiently small such that $\beta(1 - (1 - \delta)\rho) < 1$ which leads to, on the event $|\mathscr{C}_1| \geq (1 - \delta)\rho N$, exploration of the other components in $[N] \setminus H_\epsilon$ being dominated by subcritical Poisson-Galton-Watson trees which have maximal size $O_{\mathbb{P}}(\log N)$. Therefore, on this first conditioning $|\mathscr{C}_1| \geq (1 - \delta)\rho N$, the further conditioning that they do not form a larger component than $|\mathscr{C}_1|$ is asymptotically the whole probability space and has no effect. Thus we can work with independent edges outside $\mathscr{C}_1$ in the remainder of the proof (and more, by another appeal to asymptotic equivalence, return to generating these edges with the Poisson probabilities $1 - \exp(-\beta/N)$).

The next extension to the subgraph is to attach the set H_ϵ. Note between any $v \in \mathscr{C}_1$ and $h \in H_\epsilon$ we have at least mass $\beta\left(\epsilon^{-\gamma} \log N\right)$.

H is attached to $\mathscr{C}_1$ vertex by vertex: construct a function $n : H \to \mathscr{C}_1$ iteratively by giving the lowest unpaired index $h \in H_\epsilon$ its lowest neighbour in $\mathscr{C}_1 \setminus n\left([h - 1] \cap H_\epsilon\right)$ which we set as $n(h)$, and connect h to $\mathscr{C}_1$ by this single edge. Because $|H_\epsilon| < \delta\rho N$

deterministically and $|\mathscr{C}_1| > (1-\delta)\rho N$ with high probability, we have (whp) at least $(1-2\delta)\rho N$ available vertices in $\mathscr{C}_1$ to pair to. The mean number of edges to available vertices that each $h \in H$ will see is thus at least

$$(1-2\delta)\rho N \cdot \frac{1}{N}\beta\epsilon^{-\gamma}\log N > \beta \log N$$

by taking ϵ small enough such that $(1-2\delta)\rho\epsilon^{-\gamma} > 1$.

We observe $\mathbb{P}\left(\mathrm{Pois}(\beta \log N) = 0\right) = N^{-\beta} = o\left(\frac{1}{N}\right)$ and conclude by the union bound that with high probability we are successful in constructing this injective function $n : H \to \mathscr{C}_1$. The graph $\mathscr{G}_2$ is $\mathscr{G}_1$ with every vertex in H_ϵ attached as a leaf in this way, on the vertex set $\mathscr{C}_2 = \mathscr{C}_1 \cup H_\epsilon$.

Then on this same vertex set $\mathscr{C}_2$ we realise the remaining Poisson kernel to make this the *induced* subgraph on that set. Concretely, this means that between any pairs of either the form $v, w \in \mathscr{C}_1$ or the form $v \in H_\epsilon$,

$$w \in n\left([v-1] \cap H_\epsilon\right) \cup \left(\mathscr{C}_1 \setminus [n(v)]\right) \cup H_\epsilon,$$

we have a number of edges with Poisson mean

$$\frac{\beta}{N}\left(N^{2\gamma}v^{-\gamma}w^{-\gamma} - 1\right).$$

Flatten multiple edges to simplify the graph and call the resultant simple graph $\mathscr{G}_3$.

At this point we should begin to discuss the mixing times. We start with Lemma 6.1.3 which lower bounds the Cheeger constant of $\mathscr{G}_1$.

Note that a set of minimal Cheeger constant for $\mathscr{G}_2$ is simply a connected subset of $\mathscr{C}_1$ with or without pendant leaves included. Including a pendant edges will not increase $e\left(S : S^C\right)$ but will increase $|S|$ by 1, and so in fact it should include all its pendant edges. Because each vertex in H is attached to a distinct vertex in $\mathscr{C}_1$, the worst case is that every one gets a pendant edge and hence $\Phi(\mathscr{G}_2) \geq \Phi(\mathscr{G}_1)/2$.

By Theorem 6.1.2 we deduce

$$t_{\mathrm{rel}}\left(\mathscr{G}_2\right) \leq \frac{8}{\Phi^2\left(\mathscr{G}_2\right)} \max_{v \in \mathscr{C}_2} \mathrm{d}_{\mathscr{G}_2}(v) \leq \frac{32}{\Phi^2\left(\mathscr{G}_1\right)} \max_{v \in \mathscr{C}_1} (1 + \mathrm{d}_{\mathscr{G}_1}(v)) = O_{\mathbb{P}}^{\log N}\left(\log^5 N\right).$$

For the third graph $\mathscr{G}_3$ we claim by [Aldous and Fill, 2002, Corollary 3.28] that adding the internal edges did not increase the relaxation time, and so we have the same bound $t_{\mathrm{rel}}(\mathscr{G}_3) \leq t_{\mathrm{rel}}(\mathscr{G}_2)$. By applying Lemma 4.1.3 we can turn this into a bound on

the mixing time

$$t_{\text{mix}}(\mathscr{G}_3) \leq 5t_{\text{rel}}(\mathscr{G}_3)\left(1+\frac{1}{2}\log|\mathscr{C}_3|\right) = O_{\mathbb{P}}^{\log N}\left(\log^6 N\right).$$

Remember this mixing bound for the later supergraphs of $\mathscr{G}_3$. First we build $\mathscr{G}_4$ by adding in a select number of the neighbours of $\mathscr{C}_1 \cap [\lfloor N/2 \rfloor]$ which are not yet in the graph. Precisely, each vertex pair $Nx \in \mathscr{C}_1 \cap [\lfloor N/2 \rfloor]$ and $Ny \notin \mathscr{C}_3$ has remaining kernel

$$\beta\left(x^{-\gamma}y^{-\gamma}-1\right) \geq \beta\left(2^{\gamma}y^{-\gamma}-1\right) \geq \beta\left(2^{\gamma}-1\right)y^{-\gamma} \tag{6.2}$$

so we can generate this kernel by attaching $\text{Pois}\left(\beta\left(2^{\gamma}-1\right)\right)$ edges to each vertex in $\mathscr{C}_1 \cap [\lfloor N/2 \rfloor]$ and then sequentially giving each edge an i.i.d. target label from the mark distribution, or thinning it if that label is already in the graph. At the end of this process we have the graph $\mathscr{G}_4$ and its vertex set $\mathscr{C}_4$.

In the continuing construction, $\mathscr{G}_5$, we still want to not create any additional cycles – we want to span the rest of the giant component only by growing trees. Therefore we will thin any label in $\mathscr{C}_4$ in making this exploration, and any labels previously seen according to the breadth-first ordering from the initial root that found the giant (*any* ordering would work).

First, we complete the neighbourhoods of vertices in $\mathscr{C}_1$ as these have different remaining kernels seen in (6.2). Then the exploration is completed by the standard algorithm of Section 3.1. This is possible as revealed, unexplored vertices (with label, without degree) at this point, which are neighbours of $\mathscr{C}_1$ (including all of H), can only have had incident edges checked that lead to a revealed vertex (label seen in the construction). Hence we can explore as with the unmodified kernel $\beta x^{-\gamma}y^{-\gamma}$ while thinning revealed vertices.

In revealing a vertex label v we forbid it from the future exploration and reduce the offspring mean of the exploration by a certain amount $W(v)$. This has the expression

$$W(v) = \mathbb{P}(M=v)w(v) = \frac{v^{-\gamma}}{\sum_{k=1}^{N}k^{-\gamma}}\sum_{k=1}^{N}\beta N^{2\gamma-1}v^{-\gamma}k^{-\gamma} = \frac{\beta}{N}\left(\frac{N}{v}\right)^{2\gamma}$$

and so $\sum_{v\notin\mathscr{C}_4}W(v)$ is the number of unrevealed vertices that we expect to find in revealing (labelling) and then exploring (generating a degree) at a given unrevealed vertex. Technically this single type Galton-Watson formulation of the graph is giving thinned vertices zero weight rather than actually pruning them but thus it can bound the exploration, with some erroneous leaves attached.

We will describe the distribution of this upper bound for clarity. Generate a random label $M \in [N]$ with $\mathbb{P}(M = k) \propto k^{-\gamma}$, the effective weight is then only positive if $M \notin \mathscr{C}_4$ and so is the following function of M

$$W(M) = \begin{cases} \beta N^{2\gamma-1} M^{-\gamma} \sum_{k=1}^{N} k^{-\gamma} & M \notin \mathscr{C}_4 \\ 0 & M \in \mathscr{C}_4. \end{cases}$$

Offspring have the mixed Poisson distribution $D \sim \text{Pois}\,(W)$.

For every branch we explore for $\mathscr{G}_5$ (ignoring the inclusion of the ball around $\mathscr{C}_1$ of radius 1, which will add an extra edge to those trees) we can contain the real exploration of each branch, by stochastic domination, by an i.i.d. mixed Poisson Galton-Watson tree $(T_i)_i$ with offspring distribution defined by the available weight set $\mathscr{C}_4^c$. If we can argue that $\mathbb{E}(D|\mathscr{C}_4) = \mathbb{E}(W(M)|\mathscr{C}_4) = \sum_{v \notin \mathscr{C}_4} W(v) < 1$, then all the exploration in the construction of $\mathscr{G}_5$ will be contained in subcritical Galton-Watson trees.

In Lemma 6.1.6 after this proof we find $m < 1$ such that with high probability $\mathbb{E}(D|\mathscr{C}_4) < m$ when β is sufficiently large, which will make this analysis possible. To each pendant tree T_i we associate an exploration walk $X_i : \mathbb{N}^+ \to \mathbb{N}$ as in [Alon and Spencer, 2016] which represents the number of unexplored half-edges in the tree if we explore in the breadth-first ordering. Thus $X_1 = 1$, increments have the i.i.d. distribution

$$\forall t \geq 1, \quad X_i(t+1) - X_i(t) \overset{\text{(d)}}{=} D - 1$$

and the first hitting time of 0 is the size of the tree, $X_i(|T_i|) = 0$. Therefore if we fix some large constant $C > 0$ and $L = \lceil C \log^3 N \rceil$, and write $W_i \overset{\text{i.i.d.}}{\sim} W$

$$\begin{aligned} \mathbb{P}\left(|T_i| \geq L\right) &\leq \mathbb{P}\left(X_i\left(L\right) \geq 0\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{L} \text{Pois}\left(W_i\right) \geq L\right) = \mathbb{P}\left(\text{Pois}\left(\sum_{i=1}^{L} W_i\right) \geq L\right) \\ &\leq \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} W_i \geq \frac{1+m}{2}\right) + \mathbb{P}\left(\text{Pois}\left(\left(\frac{1+m}{2}\right)L\right) \geq L\right). \end{aligned}$$

Note that the second term is a Poisson large deviation and so has exponential decay in L. For the first, we observe that W is a bounded random variable

$$W \in \left[0, \frac{\beta}{1-\gamma} \log N\right] \quad \text{a.s.}$$

which will allow us to use Hoeffding's inequality [Hoeffding, 1963].

$$\mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} W_i \geq \frac{1+m}{2}\right) = \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} W_i - m \geq \frac{1-m}{2}\right)$$

$$\leq \exp\left(-\frac{2L^2\left(\frac{1-m}{2}\right)^2}{L\left(\frac{\beta}{1-\gamma}\log N\right)^2}\right) \leq \exp\left(-\frac{2C\left(\frac{1-m}{2}\right)^2}{\left(\frac{\beta}{1-\gamma}\right)^2}\log N\right)$$

So if we set C large enough so that

$$\frac{2C\left(\frac{1-m}{2}\right)^2}{\left(\frac{\beta}{1-\gamma}\right)^2} > 1$$

then no exploration will see L vertices with high probability, by the union bound.

For the VSRW on this graph $\mathscr{G}_5$ the maximal expected time to escape a pendant tree, which is maximally of size $O_{\mathbb{P}}\left(\log^3 N\right)$, is $O_{\mathbb{P}}\left(\log^6 N\right)$ by Proposition 4.1.1.

By Lemma 6.1.4 for the small weights, and from an easy argument with Poisson large deviations and the union bound for the others (which we partially did already in Proposition 3.2.3 (a)),

$$\max_{v\in[N]} \frac{\mathrm{d}(v)}{\left(\frac{N}{v}\right)^{\gamma}} = O_{\mathbb{P}}(\log N) \tag{6.3}$$

which we combine with the observation $\mathrm{d}_{\mathscr{G}_5}(v) \leq \mathrm{d}(v)$, the small ϵ assumption $\beta\rho\epsilon^{-\gamma} > 32$ and Lemma 6.1.7 after this proof to deduce that vertices in H have a bounded proportion of neighbours in $\mathscr{C}_1$

$$\min_{v\in H_\epsilon} \frac{r(v)}{\mathrm{d}_{\mathscr{G}_5}(v)} = \Omega_{\mathbb{P}}\left(\frac{1}{\log^2 N}\right)$$

so that by taking geometrically many attempts to escape vertices in H_ϵ we conclude the maximal expected time to hit the set $\mathscr{C}_1$ is $O_{\mathbb{P}}\left(\log^8 N\right)$.

The partially observed (see the definition before Theorem 4.1.10), on $\mathscr{C}_3$, version of this Markov chain is the same as the restricted to $\mathscr{C}_3$ version which we recall had mixing time $O_{\mathbb{P}}\left(\log^6 N\right)$. By [Fill, 1991] we can construct a strong stationary time T for the partially observed chain on this order, on the partially observed clock which only runs when the walker is in $\mathscr{C}_3$. We expect to hit $\mathscr{C}_1$ in maximal time $O_{\mathbb{P}}\left(\log^8 N\right)$, and by (6.3) we expect to stay in that vertex for at least time $\Omega_{\mathbb{P}}\left(\log^{-2} N\right)$. A time T with expected required occupancy time $O_{\mathbb{P}}\left(\log^6 N\right)$ expects to need at most $O_{\mathbb{P}}\left(\log^8 N\right)$ such

exponential waiting times. So, by inserting between every waiting time in $\mathscr{C}_1$ a hitting time to bring the walker back to $\mathscr{C}_1$ we conclude $\mathbb{E}(T) = O_{\mathbb{P}}\left(\log^{16} N\right)$.

After this random time T we consider the flow of probability mass. The mass $p_t(v)$ is constant on $v \in \mathscr{C}_3$ for all future times, so we must have $\frac{\mathrm{d}}{\mathrm{d}t}p_t(v)$ also constant on $v \in \mathscr{C}_3$. However because with high probability some vertices $v \in \mathscr{C}_3$ gained no edges in $\mathscr{G}_5$ we can deduce with with probability that in fact $\frac{\mathrm{d}}{\mathrm{d}t}p_t(v) = 0$ for $v \in \mathscr{C}_3$ and so T was a hitting time of the proper stationary distribution of the walk on $\mathscr{G}_5$, for the set $\mathscr{C}_3$.

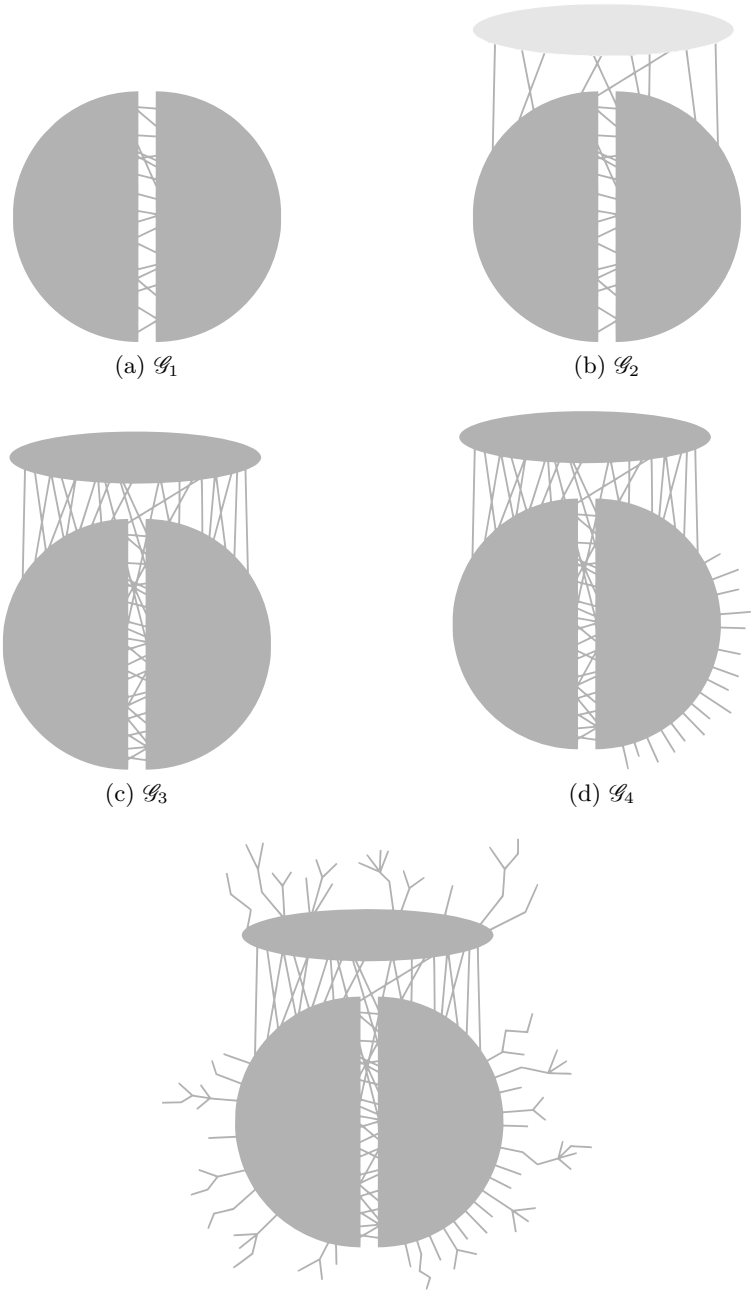
This formulation doesn't technically rule out that the pendant trees at a vertex are in disequilibrium. For a central vertex with pendant trees the maximum central hitting time is $O_{\mathbb{P}}\left(\log^6 N\right) = o_{\mathbb{P}}\left(\log^{16} N\right)$ by Proposition 4.1.1 and this bounds the mixing time on the vertex with trees by Proposition 4.1.6. The product of all of these disconnected pendant tree systems at each vertex then also has mixing time $O_{\mathbb{P}}\left(\log^6 N\right)$, and so by waiting for time $t = O_{\mathbb{P}}\left(\log^{16} N\right)$ such that $\mathbb{P}(T > t) = \frac{1}{8}$ plus the time for the trees to mix to total variation distance $\frac{1}{8}$ which is $o_{\mathbb{P}}\left(\log^{16} N\right)$ we can bound the mixing time of the VSRW on $\mathscr{G}_5$ on this order.

Finally, we complete the remaining edges internal to $\mathscr{C}_5$ which is the vertex set of the full giant component. Thus the resultant graph $\mathscr{G}_6$ is simply the SNR giant and by another application of [Aldous and Fill, 2002, Corollary 3.28] these extra edges cannot slow relaxation time, and so by Lemma 4.1.3 the final mixing time bound is on the order $O_{\mathbb{P}}\left(\log^{17} N\right)$. □

This construction $\mathscr{G}_5$ is similar in many ways to the "decorated expander" constructions of [Ding et al., 2014], although they work in discrete time.

Remark 6.1.5. *In Figure 6.2 we sketch the layers of the spanning subgraph that we construct in this proof. The first layer, $\mathscr{G}_1$, was necessary to have a fast-mixing core to build from; we aim to reach the point where the unexplored mass outside of the subgraph $\mathscr{G}$ has $\sum_{v\notin\mathscr{G}} \frac{\beta}{N}\left(\frac{N}{v}\right)^{2\gamma} < 1$ so that the final growths will be subcritical. Further to keep these trees small, we must have bounded weights (in the sense of a $\Theta^{\log N}(1)$ bound) outside $\mathscr{G}$ which we achieve in the construction $\mathscr{G}_2$. $\mathscr{G}_3$ is a minor step to keep track of mixing, and then in $\mathscr{G}_4$ we use the large β assumption to check that we have definitely reached subcritical unexplored mass. The graph $\mathscr{G}_5$ adds the promised subcritical trees and, afterwards, $\mathscr{G}_6$ completes the internal edges and is simply the full SNR giant component.*

Lemma 6.1.6. *In the above construction (given β sufficiently large) we have some*



(a) $\mathscr{G}_1$ (b) $\mathscr{G}_2$

(c) $\mathscr{G}_3$ (d) $\mathscr{G}_4$

(e) $\mathscr{G}_5$

Figure 6.2: A sketch of the first 5 subgraphs that we construct in this proof.

$m > 0$ *such that whp*

$$\sum_{v \notin \mathscr{C}_4} W(v) < m < 1.$$

Proof. Think of the network exploration in continuous time such that each vertex $i \in [N]$ is revealed to be the next unrevealed vertex (in the breadth-first order) as a Poisson process of rate $(N/i)^\gamma$. Thus all times after the first for each Poisson process will represent thinned vertices in the tree construction. Note also that the total exploration rate is

$$\sum_{i=1}^{N} \left(\frac{N}{i}\right)^\gamma \sim \frac{N}{1-\gamma}$$

By symmetry (all vertices in $[N] \setminus H_\epsilon$ are the same in the construction of $\mathscr{C}_1$) we have

$$\frac{1}{N} |\mathscr{C}_1 \cap [\lfloor N/2 \rfloor]| \xrightarrow{\mathbb{P}} \frac{\rho}{2}$$

and recall that each vertex in $\mathscr{C}_1 \cap [\lfloor N/2 \rfloor]$ explored $\mathrm{Pois}\left(\beta\left(2^\gamma - 1\right)\right)$ neighbours i.i.d. from the mark distribution to construct $\mathscr{G}_4$.

So we claim by the weak law concentration of a sum of i.i.d. $\mathrm{Pois}\left(\beta\left(2^\gamma - 1\right)\right)$ random variables, and also the concentration of a Poisson process of rate $\frac{N}{1-\gamma}$, that if

$$T := \frac{\rho}{2}(2^\gamma - 1)(1-\gamma)$$

then if we fix any $\epsilon \in (0, T)$ the continuous time exploration will have seen fewer labels than the the $\mathscr{G}_4$ exploration at time $\beta\epsilon$, with high probability.

We can therefore whp upper bound the remaining mass by thinning from the full set $[N]$ in continuous time for time $\beta\epsilon$, and in the event that the continuous exploration takes more than time $\beta\epsilon$ this is an upper bound by stochastic domination. The mass in this upper bound is the random variable

$$\mathfrak{M} = \sum_{v=1}^{N} \frac{\beta}{N} \left(\frac{N}{v}\right)^{2\gamma} \mathbb{1}_{E_v > 1}$$

where each E_v is an independent exponential of rate $\beta\epsilon\left(N/v\right)^\gamma$. We can bound this

variable with the second moment method, and so we calculate

$$\begin{aligned}\mathrm{Var}\left(\mathfrak{M}\right) &= \sum_{v=1}^{N} \frac{\beta^2}{N^2}\left(\frac{N}{v}\right)^{4\gamma}\left(1-e^{-\beta\epsilon\left(\frac{N}{v}\right)^{\gamma}}\right)e^{-\beta\epsilon\left(\frac{N}{v}\right)^{\gamma}} \\ &\leq \sum_{v=1}^{N} \frac{\beta^2}{N^2}\left(\frac{N}{v}\right)^{4\gamma} e^{-\beta\epsilon\left(\frac{N}{v}\right)^{\gamma}} \sim \frac{1}{N}\int_0^1 \beta^2 x^{-4\gamma} e^{-\beta\epsilon x^{-\gamma}}\mathrm{d}x \to 0\end{aligned}$$

as $N \to \infty$ because the integrand is bounded. Hence, the mean $\mathfrak{M}$ converges in probability as $N \to \infty$. Furthermore

$$\mathbb{E}(\mathfrak{M}) = \sum_{v=1}^{N} \frac{\beta}{N}\left(\frac{N}{v}\right)^{2\gamma} e^{-\beta\epsilon\left(\frac{N}{v}\right)^{\gamma}} \to \beta\int_0^1 x^{-2\gamma} e^{-\beta x^{-\gamma}}\mathrm{d}x$$

so it remains to check when this limit is less that 1. The integrand is non-negative, continuously differentiable and tends to 0 at 0+ and $e^{-\beta\epsilon} \to 0$ at 1−, so we can differentiate to find a unique turning point at $x^{-\gamma} = 2/\beta\epsilon$. By inserting this into the integral we find the upper bound

$$\beta \cdot 1 \cdot \frac{4}{\beta^2\epsilon^2} e^{-2} \to 0$$

as $\beta \to \infty$. Thus for any $\gamma \in (0,1)$ and $\epsilon > 0$ we have

$$\lim_{\beta\to\infty} \beta\int_0^1 x^{-2\gamma} e^{-\beta\epsilon x^{-\gamma}}\mathrm{d}x = 0$$

and so by identifying this as the limit of $\mathfrak{M}$ we conclude

$$\mathfrak{M} \xrightarrow{\mathbb{P}} \beta\int_0^1 x^{-2\gamma} e^{-\beta\epsilon x^{-\gamma}}\mathrm{d}x < 1$$

for β sufficiently large. □

Lemma 6.1.7. *Let $r(v)$ denote the number of $\mathscr{C}_1$ neighbours*

$$r(v) = \left|\{w \in \mathscr{C}_1 : w \sim v\}\right|.$$

Then if we take ϵ small enough such that $\beta\rho\epsilon^{-\gamma} > 32$, in the above construction we have

$$\min_{v\in H_\epsilon} \frac{r(v)}{\left(\frac{N}{v}\right)^{\gamma}} = \Omega_{\mathbb{P}}\left(\frac{1}{\log N}\right).$$

Proof. Any $v \in H_\epsilon$, i.e. $v \leq \epsilon N \log^{-\frac{1}{\gamma}} N$, and any $w \in \mathscr{C}_1$ have MNR mean number of

edges at least

$$\frac{\beta}{N}\left(\frac{v}{N}\right)^{-\gamma} \leq \beta N^{\gamma-1}.$$

Working on the high probability assumption that $|\mathscr{C}_1| \geq (1-\delta)\rho N \geq \frac{\rho}{2}N$ and using that the SNR graph dominates the GRG graph (i.e. that $1-e^{-p} \geq \frac{p}{1+p}$) we claim

$$r(v) \succeq \mathrm{Bin}\left(\frac{\rho}{2}N, \frac{\frac{\beta}{N}\left(\frac{v}{N}\right)^{-\gamma}}{1+\beta N^{\gamma-1}}\right) \succeq \mathrm{Bin}\left(\frac{\rho}{2}N, \frac{\frac{\beta}{N}\left(\frac{v}{N}\right)^{-\gamma}}{2}\right).$$

Hence by the usual multiplicative Chernoff bound these edges number less than $\frac{\beta\rho}{8}\left(\frac{N}{v}\right)^{\gamma}$ with probability asymptotically bounded by

$$\exp\left(-\frac{1}{8}\cdot\frac{\beta\rho}{4}\left(\frac{N}{v}\right)^{\gamma}\right) \leq \exp\left(-\frac{\beta\rho}{32}\epsilon^{-\gamma}\log N\right) = o\left(\frac{1}{N}\right),$$

by the assumption $\beta\rho\epsilon^{-\gamma} > 32$. The conclusion follows from the union bound over every vertex in H_ϵ. □

6.2 Consequences for the Discursive Voter Model

The VSRW Markov chain is the dual of the discursive voter model with $\theta = 1$. Fortunately this induces a bound over every $\theta \geq 1$ version.

Proposition 6.2.1. *On any connected graph the discursive dual relaxation time $\theta \mapsto t_{\mathrm{rel}}^{(\theta)}$ is non-increasing.*

Proof. Recall the transition rates are

$$Q(i,j) = \frac{\mathrm{d}(i)^{\theta-1} + \mathrm{d}(j)^{\theta-1}}{2}\mathbb{1}_{i\sim j}$$

for every $i \neq j$. So, the matrix Q is symmetric, i.e. this Markov chain is a "fluid model" in the sense of [Aldous and Fill, 2002]. The claim then follows immediately from [Aldous and Fill, 2002, Corollary 3.28]. □

Proof of Theorem 2.1.13. Theorem 2.1.12 bounds the mixing time on the order $O_{\mathbb{P}}^{\log N}(1)$, and Lemma 4.1.3 gives $t_{\mathrm{rel}} \leq t_{\mathrm{mix}}$ so we have the same bound on the relaxation time. Then, Proposition 6.2.1 above makes this a uniform bound over every $\theta \geq 1$ and another application of Lemma 4.1.3 for the reverse direction (using that every discursive model

has uniform stationary distribution)

$$t_{\text{mix}} \leq 5t_{\text{rel}} \left(1 + \frac{1}{2} \log N\right)$$

translates this to a bound of the mixing time for any $\theta \geq 1$ model.

We can relate the mixing time to the probability of being at the initial state via [Aldous and Fill, 2002, (3.61)]. Writing π for the unique stationary distribution on each component, the symmetric generator for this model gives

$$\pi(v) = \frac{1}{|\mathscr{C}(v)|}$$

and hence, writing $p_{v,\cdot}^{(t)}$ for the kernel of the discursive dual, we have

$$p_{v,v}^{(2t)} - \frac{1}{|\mathscr{C}(i)|} = \left\| p_{v,\cdot}^{(t)} - \frac{1}{|\mathscr{C}(i)|} \right\|_2^2 \leq \left\| p_{v,\cdot}^{(t)} - \frac{1}{|\mathscr{C}(i)|} \right\|_1 \max_k \left| p_{k,\cdot}^{(t)} - \frac{1}{|\mathscr{C}(k)|} \right| \leq 2d(t)$$

by recalling that the total variation distance is half the L^1 distance. Then by [Levin et al., 2009, Lemma 4.11] we have $\bar{d}(t) \leq 2d(t)$ and so $\bar{d}(t_{\text{mix}}) \leq \frac{2}{e}$. Because $\bar{d}$ is submultiplicative, we say further

$$\bar{d}\left(t_{\text{mix}} \lceil C \log N \rceil\right) \leq N^{-C \log\left(\frac{e}{2}\right)} = \frac{1}{N}$$

by setting $C = \frac{1}{1 - \log 2}$. The reverse bound of [Levin et al., 2009, Lemma 4.11], $d(t) \leq \bar{d}(t)$, makes this a bound on total variation distance. Then, the covariance is straightforward by recalling $\mathbb{E}\left(\eta_0(v)\right) = \mathbb{E}\left(\eta_t(v)\right) = u$ and

$$\mathbb{E}\left(\eta_0(v)\eta_t(v)\right) = \mathbb{P}\left(\eta_0(v) = \eta_t(v) = 1\right) = up_{v,v}^{(t)} + u^2\left(1 - p_{v,v}^{(t)}\right)$$

so that

$$\text{Cov}\left(\eta_0(v), \eta_t(v)\right) = p_{v,v}^{(t)} u(1-u)$$

and we do not need to avoid the extra logarithmic factor in proving the claim for times $\omega^{\log N}(1)$.

For the other components we have a similar argument using that by [Bollobás et al., 2007, Theorem 3.12(ii)] they can be uniformly bounded at size $O_{\mathbb{P}}(\log N)$ and so by Proposition 4.1.1 the hitting time of the chain on the maximal small component is $O_{\mathbb{P}}(\log^2 N)$. By Lemma 4.1.5 we see the mixing time must also be $O_{\mathbb{P}}(\log^2 N)$. □

For the consensus time in this regime we have a partial result Theorem 2.1.10, though as claimed in the discussion of this result the region $\theta \leq \frac{1}{\gamma}$ covers the likely modelling parameters.

Proof of Theorem 2.1.10. With high probability, Lemma 3.3.5 gives a set of paths connecting every pair in $S := \left[\left\lfloor N \log^{-2/\gamma} N \right\rfloor\right]$ by paths in $[N] \setminus S$, of maximal length $O_{\mathbb{P}}(\log N)$.

The diameter result in Theorem 3.3.4 gives a set of paths between every pair of vertices in $\mathscr{C}(1)$, of maximal length $O_{\mathbb{P}}(\log N)$.

Therefore we can alter paths in the second set which have more than 2 vertices in S, by taking the first and last vertex in S, deleting the path between them and replacing it by their low degree path from the first set.

We use that

$$\max_{v > N \log^{-2/\gamma} N} \mathrm{d}(v) = O_{\mathbb{P}}^{\log N}(1)$$

to say that the conductance of any edge $\{v, w\}$ incident to a vertex $v \notin S$ has

$$c(v,w) \geq \frac{1}{|\mathscr{C}(1)|} \frac{\mathrm{d}(v)^{\theta}}{2} \geq \frac{1}{N} \frac{\mathrm{d}(v)^{\theta}}{2}$$

to uniformly bound the expected hitting of each pair on the order $O_{\mathbb{P}}^{\log N}(N)$ via the electrical network bound 4.1.1. This induces a bound on the expected consensus time by Proposition 4.2.5.

Further we have an upper bound on the small components, which by [Bollobás et al., 2007, Theorem 3.12(ii)] have maximal size $O_{\mathbb{P}}(\log N)$, by the same hitting time argument – using that the maximum degree is bounded by the maximum component size.

The lower bound follows by an application of Proposition 4.1.4 (a) to $\mathscr{C}(1)$. This simplifies, because the stationary distribution is uniform on $\mathscr{C}(1)$, to $\Omega_{\mathbb{P}}(N^2 / \sum_v q(v))$. Whenever

$$\sum_{v \in \mathscr{C}(1)} q(v) = \sum_{v \in \mathscr{C}(1)} \mathrm{d}(v)^{\theta} = O_{\mathbb{P}}^{\log N}\left(\sum_{v=1}^{N} \left(\frac{N}{v} \right)^{\gamma} \right) = O_{\mathbb{P}}^{\log N}(N)$$

we have a lower bound which is polylogarithmically tight to the upper bound, and this happens when $\theta \leq \frac{1}{\gamma}$. □

Chapter 7

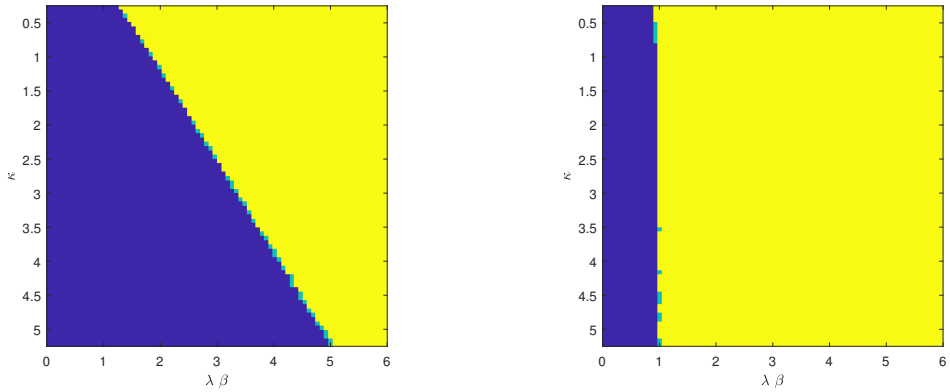
The Contact Process on a Network Adapting to Infection Density

In this chapter we prove Theorem 2.2.8 on when the epidemic event breaking out from a single source has asymptotic probability zero, and 2.2.10 on when it has asymptotically positive probability. In both cases we look to find an explicit region in our three parameters (λ, κ, β) controlling the relative rate of infection, relative rate of the dynamic and (initial) mean degree. Results for the large graph limit $N \to \infty$ might then help us understand behaviour on the finite but large networks that real epidemics occupy.

Not as a main result but for context, we first consider the non-adaptive version of the graph dynamic where every vertex updates its neighbourhood at constant rate regardless of its neighbourhoods infection level. With this dynamic, the random graph is independent of the infection and hence we are able to apply some existing techniques, imitating [Jacob and Mörters, 2017], to prove the results of Theorem 2.2.11.

In the second part of the chapter we modify the dynamic such that the neighbourhood update of a vertex is only carried out if there is an adjacent infected vertex. This graph dynamic for an infection model is one of the most natural from a social dynamics perspective, and is superficially similar to the non-adaptive version.

However, the modification has broken many useful mathematical features that were previously present. As well as losing self-duality for the contact process on this adaptive dynamic graph, we lose all types of monotonicity: an extra edge on the graph could cause an update and thus hinder the spread of infection; and an extra infected vertex could in the same way lead to a reduction in the resultant size of the infection set



(a) Adaptive (b) Non-adaptive

Figure 7.1: Recalling the simulations previously shown in Figures 2.2 and 2.3 but now on common axes. Each is with $\mu = 10$, and requiring two standard deviations of confidence after 10^5 samples to colour a pixel yellow (for epidemic probability zero) or blue (for positive epidemic probability).

I_∞. Furthermore, the graph distribution is no longer stationary, firstly because the empty graph is an absorbing state for the dynamic. More significantly, while vertices are updated with the same (approximately) Pois(β) distribution, over time we will see the process spend more time in lower degree states because lower degree vertices are less likely to see an infected neighbour.

Recently [Ball et al., 2019] have considered a similarly dependent graph model but with only edge removal and no replacement. Also, they consider the SIR dynamic where recovery is permanent, which allows them to eliminate some dependence in the local picture of the spreading infection. We keep the full complication of the SIS model but in exchange we drop the heterogeneity that was dealt with for the voter model, effectively fixing $\gamma = 0$. In the homogenous context we will not encounter the issues of [Pastor-Satorras and Vespignani, 2002] with polynomially large degrees, but we still cannot use the mean-field approximation because the infection has become significantly dependent with the network.

We show that the contact process on this adaptive dynamic graph cannot grow to epidemic levels by constructing a stochastic upper bound which is more tractable and also cannot grow to epidemic levels. Because the model is substantially difficult to work with, we must make substantial modifications and so we arrive at a loose upper bound model. This model we call the subtree contact process (SCP), and it bounds the contact process on a rooted tree by disallowing all recoveries that would disconnect the infection

set, thus constructing a *reversible* Markov process.

For the expected size of the infection set in this bounding model to not itself diverge, which would make the model useless for obtaining the theorem, $\lambda\beta < e^{-1}$ is required. From the local perspective of a vertex in the network, the original dynamic of changing neighbourhood whenever you see infection sounds like it should work to avoid epidemics, making epidemics asymptotically impossible for a larger parameter set than with non-adaptive updating. Indeed in Figure 7.1 we see this supported and that very likely the whole region $\lambda\beta < 1$ is included. Unfortunately, without any tighter upper bound we cannot hope to prove this claim.

We apply this bound to every local neighbourhood reached by the infection, using the intuition that local neighbourhoods are treelike. Further because the typical distance for the Erdős-Rényi giant component is $\Theta_{\mathbb{P}}(\log N)$, updates even if they keep a vertex connected to the same component are expected to connect it far away. Thus we imagine the initial phase of the epidemic as occupying a dynamic forest model which we formalise as the "contact process on an evolving forest" (CPEF), which could be thought of as the $N = \infty$ version of the infection model. By working with the SCP bound in the CPEF picture, we show that epidemics are asymptotically impossible whenever $\lambda\beta < 0.21$.

In the section on infection supercriticality we will instead find a stochastic lower bound which grows to epidemic levels. For this we use an SIR coupling which shows survival of the infection just on the root tree of the CPEF. This might be called *tree supercriticality*, and it requires considerable higher λ parameter than for the real object of interest which is *forest supercriticality*: possible survival on the whole CPEF regardless of possible survival on a single tree. In Figure 2.2 we can see two lines of very different slopes which bound these two regions, but we have no conjecture for where the line falls between forest supercriticality and subcriticality. We do, however, conjecture that this is the same line that bounds the maximal region of asymptotic epidemic possibility on the finite network, and visibly it contains the region proved by this SIR lower bound.

7.1 The Non-adaptive Dynamic

First, as in [Jacob and Mörters, 2017, Jacob et al., 2019] we consider an independent network dynamic which is invariant to time reversal. So, this context is amenable to existing techniques which we will apply in this section.

Proof of Theorem 2.2.11 (a). We use the mean-field upper bound of [Jacob and Mörters, 2017] $(Y_t)_t$ which has for each vertex v three states $Y_t(v) \in \{0, 1, 2\}$. Of these, 0 means

healthy, 2 means *infected* and the intermediate state 1 means *ready to recover*. In this model we do not keep track of a dynamic graph (because it is instead mean-field) but have the following dynamic: every infected vertex "updates" at rate κ and becomes ready to recover, and every ready to recover vertex recovers at rate 1 and becomes healthy.

Both infected and ready to recover vertices infect every other vertex at rate $\lambda\beta/N$, after which (if only ready to recover) the infecter also becomes infected. Thus after an infection event both involved vertices become fully infected at value 2. They show that this model can be coupled to the contact process ξ and the non-adaptive dynamic graph such that $\xi \leq Y$.

This mean-field upper bound is designed for a supermartingale analysis which is simpler in the homogeneous case. However because we are not looking to take λ arbitrarily small we should optimise the relative value of $Y = 2$ and $Y = 1$ which in the following martingale,

$$M(t) := \#\{v : Y_t(v) = 1\} + C\#\{v : Y_t(v) = 2\},$$

is controlled by $C \geq 1$. Recall the dynamic of Y to bound

$$\begin{aligned}\frac{1}{\mathrm{d}t}\mathbb{E}\left(M(t+\mathrm{d}t) - M(t)\middle|\mathscr{F}_t\right) \leq{}& \left(\lambda\beta(2C-1) - 1\right)\#\{v : Y_t(v) = 1\} \\ &+ \left(\lambda\beta C + \kappa(1-C)\right)\#\{v : Y_t(v) = 2\}\end{aligned}$$

and we can minimise the maximum of the two bracketed coefficients over C by setting

$$\lambda\beta(2C-1) - 1 = \frac{\lambda\beta C + \kappa(1-C)}{C}$$

which happens at

$$C = \frac{1 - \kappa + 2\lambda\beta + \sqrt{(1-\kappa)^2 + 4\lambda\beta(1+\kappa+\lambda\beta)}}{4\lambda\beta}.$$

This leads to a bound of the form

$$\frac{1}{\mathrm{d}t}\mathbb{E}\left(M(t+\mathrm{d}t) - M(t)\middle|\mathscr{F}_t\right) \leq -\epsilon M(t)$$

with $\epsilon > 0$ possible if and only if $\lambda\beta < 1$ and

$$\kappa > \frac{\lambda\beta(1+\lambda\beta)}{1-\lambda\beta}.$$

Given parameters satisfying these requirements, we can apply optional stopping to

the supermartingale $\left(M(t)e^{\epsilon t}\right)_t$ to deduce

$$\mathbb{E}_{[N]}\left(\sum_{v=1}^{N}\xi_t(v)\right) \leq \mathbb{E}\left(M(t)\right) \leq M(0)e^{-\epsilon t} = Ne^{-\epsilon t}$$

by setting $M(0) \equiv 1$ so everywhere is initially in the ready to recover state.

There is no immediate link to the main quantity of interest $\mathbb{E}_1(|I_\infty|)$ but we can observe that for the recovery time R

$$\mathbb{E}_1\left(R\middle||I_\infty|\right) \geq \sum_{k=1}^{|I_\infty|}\frac{1}{k} \geq \log|I_\infty|$$

and hence from the duality result in Corollary 4.3.4

$$\mathbb{E}_1\left(\log|I_\infty|\right) \leq \mathbb{E}_1\left(R\right) = \mathbb{E}_{[N]}\left(\int_0^\infty \frac{1}{N}\sum_{v=1}^{N}\xi_t(v)\mathrm{d}t\right) \leq \int_0^\infty e^{-\epsilon t}\mathrm{d}t = O(1)$$

which proves tightness of $(|I_\infty|)_N$, implying that for any $\epsilon > 0$ the epidemic event E_ϵ^N does not occur, with high probability. □

To show survival with the non-adaptive dynamic as in Theorem 2.2.11 (b) we will lower bound the growth by a Pólya urn.

Definition 7.1.1 (Hazard rate)**.** A random variable X on $[0,\infty)$ which is absolutely continuous with respect to the Lebesgue measure has hazard rate function

$$q(x) := \lim_{\delta\downarrow 0}\frac{1}{\delta}\mathbb{P}\left(X \in [x, x+\delta]\middle|X \geq x\right).$$

In the calculations for bounding this rate we use the well known result that for a random variable X on $\mathbb{N}$ we have $\mathbb{E}(X) = \sum_{x=1}^{\infty}\mathbb{P}(X \geq x)$. Further, we need the following simple result.

Proposition 7.1.2. *For a random variable X on $\mathbb{N}$ we have*

$$\sum_{x=1}^{\infty} x\mathbb{P}(X \geq x) = \frac{\mathbb{E}(X) + \mathbb{E}(X^2)}{2}.$$

Proof.

$$\sum_{x=1}^{\infty} x\mathbb{P}(X \geq x) = \sum_{x=1}^{\infty} x \sum_{y=x}^{\infty} \mathbb{P}(X = y) = \sum_{y=1}^{\infty} \mathbb{P}(X = y) \sum_{x=1}^{y} x = \sum_{y=1}^{\infty} \mathbb{P}(X = y) \left(\frac{y(y+1)}{2} \right).$$

□

Lemma 7.1.3. *Consider a dynamic star structure with one permanently infected vertex having outdegree* $(\mathrm{d}^+(t))_t$ *and a healthy neighbourhood. Set*

$$\mathrm{d}^+(0) \sim \mathrm{Pois}(\beta)$$

and the dynamic

$$\begin{aligned} &\mathrm{d}^+ \to \mathrm{d}^+ - 1 \text{ at rate } \kappa\, \mathrm{d}^+ \\ &\mathrm{d}^+ \to \mathrm{d}^+ + 1 \text{ at rate } \kappa\beta \\ &\mathrm{d}^+ \sim \mathrm{Pois}(\beta) \text{ at rate } \kappa \end{aligned} \tag{7.1}$$

then we find as $\kappa \to \infty$ *the hazard rate* q *for the first infection time, which arrives throughout at rate* $\lambda\, \mathrm{d}^+(t)$*, satisfies*

$$\inf_{x \geq 0} q(x) \to \lambda\beta.$$

Proof. The full Markov chain under consideration is on $\mathbb{N} \cup \{\dagger\}$, with the degree dynamic described above together with a "death" rate $\lambda\, \mathrm{d}^+$ from each state $\mathrm{d}^+ \in \mathbb{N}$, leading to the absorbing state $\dagger$ which represents the first infection time. The definition of q conditions on this transition not being made and so we are considering a *metastable* version of the chain, which has some stationary distribution α different to the stationary distribution π of the degree dynamic, $\alpha \neq \pi = \mathrm{Pois}(\beta)$, as discussed in Section 4.1.3. Recall from that section the defining equation for each $j \in \mathbb{N}$

$$\kappa \sum_{i=0}^{\infty} \alpha_i q_{ij} - \lambda j \alpha_j = -\rho \alpha_j. \tag{7.2}$$

where $Q = [q_{i,j}]$ is the $\kappa = 1$ degree dynamic. Consider some $x \in \mathbb{N}^+$ and define

$$a = [0, x-1] \cap \mathbb{Z}, \qquad b = [x, \infty) \cap \mathbb{Z}.$$

The infection rate at any state in b is at least λx, whereas in a it is at most $\lambda(x-1)$.

So we can apply Proposition 4.1.11 to these two sets and obtain

$$\kappa \sum_{i=1}^{x-1} \sum_{j=x}^{\infty} \alpha(i) q_{ij} \geq \kappa \sum_{i=1}^{x-1} \sum_{j=x}^{\infty} \alpha(j) q_{ji}.$$

Insert the degree dynamic of (7.1) to obtain a bound independent of κ

$$\alpha(a)\pi(b) + \beta\alpha(x-1) \geq \alpha(b)\pi(a) + x\alpha(x)$$

and subsequently the inequality that we will use repeatedly

$$\pi(b) + \beta\alpha(x-1) \geq \alpha(b) + x\alpha(x). \tag{7.3}$$

Let A, Π be random variables with probability mass α, π and sum (7.3) to find

$$\sum_{x=1}^{\infty} (\pi(b) + \beta\alpha(x-1)) = \mathbb{E}(\Pi) + \beta \geq \sum_{x=1}^{\infty} (\alpha(b) + x\alpha(x)) = 2\mathbb{E}(A)$$

and so by recalling $\pi = \text{Pois}(\beta)$ we have $\mathbb{E}(A) \leq \beta$. Using Proposition 7.1.2 we can also bound the second moment from (7.3):

$$\begin{aligned}
&\sum_{x=1}^{\infty} x\,(\pi(b) + \beta\alpha(x-1)) = \frac{\mathbb{E}(\Pi) + \mathbb{E}(\Pi^2)}{2} + \beta(1 + \mathbb{E}(A)) = 2\beta + \frac{\beta^2}{2} + \beta\mathbb{E}(A) \\
&\geq \sum_{x=1}^{\infty} x\,(\alpha(b) + x\alpha(x)) = \frac{\mathbb{E}(A) + \mathbb{E}(A^2)}{2} + \mathbb{E}(A^2) = \frac{1}{2}\mathbb{E}(A) + \frac{3}{2}\mathbb{E}(A^2)
\end{aligned}$$

so that we conclude $\mathbb{E}(A^2) \leq \frac{4}{3}\beta + \frac{\beta^2}{3} + \frac{2}{3}\beta\mathbb{E}(A) \leq \frac{4}{3}\beta + \beta^2$.

The third and final application of (7.3) is to check that $(\alpha)_{\kappa>0}$ is uniformly integrable, which follows shortly. Take $L \in \mathbb{N}$; we first observe for a random variable X on $\mathbb{N}$ that

$$\mathbb{E}(X; X \geq L) = L\mathbb{P}(X \geq L) + \mathbb{E}\left((X-L)^+\right) = L\mathbb{P}(X \geq L) + \sum_{x=L+1}^{\infty} \mathbb{P}(X \geq x)$$

then by (7.3)

$$\begin{aligned}
&\mathbb{E}(\Pi; \Pi \geq L) - L\pi([L,\infty)) + \beta\alpha([L,\infty)) \\
&= \sum_{x=L+1}^{\infty} (\pi(b) + \beta\alpha(x-1)) \\
&\geq \sum_{x=L+1}^{\infty} (\alpha(b) + x\alpha(x)) \\
&= \mathbb{E}(A; A \geq L) - L\alpha([L,\infty)) + \mathbb{E}(A; A \geq L+1) \\
&= 2\mathbb{E}(A; A \geq L) - L\alpha(L) - L\alpha([L,\infty))
\end{aligned}$$

which on rearrangement gives the bound

$$\mathbb{E}(A; A \geq L) \leq \frac{L\alpha(L) + L\alpha([L,\infty)) + \mathbb{E}(\Pi; \Pi \geq L) - L\pi([L,\infty)) + \beta\alpha([L,\infty))}{2}.$$

By Markov's inequality applied to A^2 we have

$$\sup_{\kappa>0} L\alpha([L,\infty)) \leq \frac{\mathbb{E}(A^2)}{L} \leq \frac{\frac{4}{3}\beta + \beta^2}{L} \to 0$$

and so because Π is also integrable we see that every term tends to zero uniformly and

$$\sup_{\kappa>0} \mathbb{E}(A; A \geq L) \to 0;$$

we will use all three bounds in the remaining proof.

By summing over j in (7.2) we obtain $\beta \geq \mathbb{E}(A) = \rho/\lambda$ and therefore, again using (7.2), we can bound the 2 norm

$$\begin{aligned}
\|\alpha Q\|_2^2 = \sum_{j=0}^{\infty} \left(\frac{\lambda j - \rho}{\kappa}\right)^2 \alpha_j^2 &\leq \frac{\lambda^2}{\kappa^2} \sum_{j=0}^{\infty} j^2 \alpha_j^2 + \frac{\rho^2}{\kappa^2} \sum_{j=0}^{\infty} \alpha_j^2 \\
&\leq \frac{\lambda^2}{\kappa^2} \mathbb{E}(A^2) + \frac{\rho^2}{\kappa^2} \leq \frac{\lambda^2}{\kappa^2} \left(\frac{4}{3}\beta + 2\beta^2\right).
\end{aligned}$$

Further, we have an easy bound on the mixing time of the degree dynamic Q via the regeneration times which arrive as a Poisson process of rate 1. At time $\log 4$ we have seen such a time with probability $\frac{1}{4}$ and hence $t_{\text{mix}} \leq \log 4$. Recalling Lemma 4.1.3 it follows that $t_{\text{rel}} \leq \log 4$ and so every eigenvalue of Q apart from the single zero eigenvalue is less than $-\frac{1}{\log 4}$. It follows that as $\kappa \to \infty$ we have $\alpha \xrightarrow{L^2} \pi$. Then by

applying [Kallenberg, 2006, Proposition 4.12] we deduce first that $\alpha \xrightarrow{\mathbb{P}} \pi$, and applying [Kallenberg, 2006, Proposition 4.12] again recalling the uniform integrability checked above we finally deduce $\alpha \xrightarrow{L^1} \pi$.

That is to say, for any $\epsilon > 0$ by taking κ sufficiently large we guarantee total variation distance $d_{\mathrm{TV}}(\mathrm{d}^+(0), \alpha) \leq \epsilon$ and so we can couple the initial and metastable stationary distributions in an event $\mathfrak{C}$ with $\mathbb{P}(\mathfrak{C}) = 1 - \epsilon$. Then, initiated at the metastable distribution, we have a precisely exponentially distributed infection time with rate ρ. Thus for $x, \delta > 0$ the first infection time T has

$$\mathbb{P}\left(T \in [x, x+\delta] \middle| T \geq x\right) \geq \mathbb{P}\left(T \in [x, x+\delta] \middle| T \geq x, \mathfrak{C}\right) \mathbb{P}(\mathfrak{C}) = \left(1 - e^{-\rho\delta}\right)(1-\epsilon)$$

and so, recalling Definition 7.1.1, q the hazard rate of T satisfies

$$q(x) = \lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{P}\left(T \in [x, x+\delta] \middle| T \geq x\right) \geq (1-\epsilon)\rho$$

and the L^1 convergence also tells us that, as $\kappa \to \infty$, $\rho = \lambda \mathbb{E}(A) \to \lambda\beta$ so we have the result, by taking $\epsilon \downarrow 0$. □

This result on the infection rate of an updating star is the main part of proving Theorem 2.2.11 (b), it essentially just remains to find these stars on the finite network.

Proof of Theorem 2.2.11 (b). By the monotonicity properties of the contact process on a non-adaptive dynamic network, as described in [Jacob and Mörters, 2017], we can show survival on the full model by showing it on a model with stochastically fewer edges. Define some maximal outdegree L, $P \sim \mathrm{Pois}(\beta(1-\epsilon))$ and

$$P_L \overset{\mathrm{d}}{=} P \big| \{P \leq L\}.$$

While the infection set $(\xi_t)_t$ has size

$$(1+L)\,|\xi_t| \leq \frac{\epsilon N}{2}$$

we can use the stochastic domination (for N sufficiently large)

$$P_L \preceq \mathrm{Bin}\left(\left\lfloor \left(1 - \frac{\epsilon}{2}\right) N \right\rfloor, \frac{\beta}{N}\right)$$

to give each infective vertex a P_L-distributed outdegree and have every such neighbour *unique* to a member of ξ_t.

We construct the following lower bound to the infection set: vertices in ξ_t are either active or dormant. Active vertices have the star dynamics of Lemma 7.1.3 but restricted to $[L]$, when looking at their healthy neighbours. As $\kappa \to \infty$, we lower bound the hazard rate for the restricted process by some limit rate $\lambda\beta_L$ where $\beta_L := \mathbb{E}(P_L) \to \beta$ as $L \to \infty$. Hence for any $\delta > 0$, for κ sufficiently large, we can lower bound this rate by $\lambda\beta_L(1-\delta)$ by using Lemma 7.1.3, with the slight modification of the upper bound L.

If an active vertex infects one of its healthy neighbours, the neighbour becomes active infected and the infector is labelled *dormant*. We ignore inactive dormant vertices until the time of their next vertex update when they again become active. So given this decomposition

$$|\xi_t| = A_t + D_t$$

we can see this as a Pólya urn contained within the process, with mean matrix (see [Janson, 2004])

$$A = \begin{pmatrix} -1 & \kappa \\ \lambda\beta_L(1-\delta) & -1-\kappa \end{pmatrix}$$

which has a positive eigenvalue for large κ if $\lambda\beta_L(1-\delta) > 1$ which is possible for large L and small δ whenever $\lambda\beta > 1$. We conclude that survival of the process is possible, by standard multitype Galton-Watson results first shown in [Harris, 1963]. □

7.2 The Adaptive Dynamic

7.2.1 Subcriticality

The only tool we have to control the spread of the infection is the "Subtree Contact Process" that we introduced in Definition 4.3.8. Fortunately, analysis of this process on the Poisson-Galton-Watson tree is very feasible when we look at the annealed context, averaging over this Poisson-Galton-Watson tree random environment.

Lemma 7.2.1. *On the random (finite or infinite)* Pois(β)*-Galton-Watson tree we have, for the normalising constant Z_λ defined in* (4.12)*,*

$$\mathbb{E}(Z_\lambda) \leq 1 + \lambda + \beta\lambda^2 + \frac{e^3}{3\sqrt{6\pi}} \cdot \frac{\lambda^3\beta^2}{1-\lambda\beta e}$$

whenever $\lambda\beta e < 1$.

Proof. For some positive integer k, let s_k denote the expected number of subtrees of size k containing the root $\circledcirc$ in a Pois(β)-Galton-Watson tree and let $s_k^{(N)}$ denote the

expected number of subtrees of size k containing the vertex 1 in the Erdős-Rényi Graph with parameter β/N.

First we remark that $s_1^{(N)} = 1$ and $s_2^{(N)} = \mathbb{E}(\mathrm{d}(1)) = \beta\left(1 - \frac{1}{N}\right) \leq \beta$. We define $s_0^N = 1$ as the SCP should have only one empty state.

For $k \geq 3$, in [Chin et al., 2018] we see (from Cayley's formula) that the complete graph has $\binom{N}{k}k^{k-2}$ subtrees of size k and so similarly it has $\binom{N}{k-1}k^{k-2}$ subtrees of size k containing the vertex 1. Each such tree is seen as a subgraph of the Erdős-Rényi Graph with probability $(\beta/N)^{k-1}$, and hence by linearity of the expectation

$$\begin{aligned}
s_k^{(N)} &= \left(\frac{\beta}{N}\right)^{k-1}\binom{N}{k-1}k^{k-2} \\
&\to \frac{\beta^{k-1}k^{k-2}}{(k-1)!} = \frac{\beta^{k-1}k^{k-1}}{k!} \qquad \text{as } N \to \infty \\
&\leq \frac{\beta^{k-1}k^{k-1}}{\sqrt{2\pi k}k^k e^{-k}} \\
&= \frac{1}{\beta\sqrt{2\pi}k^{3/2}}(\beta e)^k \\
&\leq \frac{1}{3\beta\sqrt{6\pi}}(\beta e)^k \qquad \text{because } k \geq 3.
\end{aligned}$$

Any subtree of size k is contained in the ball $B(1,k)$ and by [van der Hofstad, 2020, Theorem 2.11] we have local weak convergence in probability, which implies local weak convergence in distribution. The number of rooted subtrees of size k is not a bounded function, but the well-known "Portmanteau Lemma" still gives us

$$\lim_{N\to\infty} s_k^{(N)} \geq s_k.$$

Hence

$$\begin{aligned}
\mathbb{E}(Z_\lambda) = \sum_{k=0}^{\infty}\lambda^k s_k &\leq 1 + \lambda + \beta\lambda^2 + \frac{1}{3\beta\sqrt{6\pi}}\sum_{k=3}^{\infty}(\lambda\beta e)^k \\
&= 1 + \lambda + \beta\lambda^2 + \frac{e^3}{3\sqrt{6\pi}} \cdot \frac{\lambda^3\beta^2}{1-\lambda\beta e}.
\end{aligned}$$

□

Controlling the mean of Z_λ allows us to control the mean expected recovery time for the SCP. We want to convert this information about the recovery time to something about the total spread of the infection, so to this end we introduce a *size-slowed* version

for which bounds on the recovery time will suggest small infection set sizes.

Definition 7.2.2 (Slowed Subtree Contact Process)**.** Let the Slowed Subtree Contact Process (Slowed SCP), for some constant $\rho > 1$, be the SCP dynamic but leaving each tree state T proportionally $\rho^{-|T|}$ slower. Hence each valid recovery occurs at rate $\rho^{-|T|}$, each infection at rate $\lambda\rho^{-|T|}$ and the stationary distribution of this process is $\pi_{\lambda\rho}$, in the same family of distributions as for the SCP.

To keep finite SCP expected recovery time with the addition of this slowing parameter ρ, because the stationary measure is the same as for the infection with $\lambda \mapsto \lambda\rho$, we must have $\lambda\beta\rho e < 1$. Allowing larger ρ values gives better control over the infection size, but while maintaining $\lambda\beta\rho e < 1$ it can limit the maximal permitted value of $\lambda\beta$. By some numerical computations that we will not show, we calculate that the following assumption is very close to optimal for maximising the subcriticality region in the resultant main theorem: Theorem 2.2.8.

Lemma 7.2.3. *If $\lambda\beta e \leq \frac{3}{4}$ then the root tree infection set $I_\infty^\emptyset$, of vertices that ever saw the infection as in Definition 2.2.2, has*

$$\mathbb{E}(|I_\infty^\emptyset|) \leq 1 + \frac{27\beta\lambda}{16}\left(1 + \sqrt{\frac{2}{3\pi}} \cdot \frac{2e^3\beta\lambda}{9 - 12\beta\lambda e}\right).$$

In fact, we prove the same bound for the $\kappa = 0$ model: the contact process on a static Galton-Watson tree.

Proof. It is sufficient to completely neglect updating on each local tree set. Of course, locally, updating is simply deletion of a vertex and prevention of infection via its incident edges – therefore the infection with no updating dominates that with updating.

To analyse the spread of an infection without updating, we can look at the Slowed SCP as an upper bound. For some

$$\rho \in \left(1, \frac{1}{\beta\lambda e}\right) \supset \left(1, \frac{4}{3}\right)$$

we verify from the detailed balance equations that the distribution $\pi_{\lambda\rho}(T) = (\lambda\rho)^{|T|}/Z_{\lambda\rho}$ of Equation (4.12) is the stationary distribution, which will allow us to bound recovery times via Lemma 7.2.1.

Write $\mathcal{T}$ for the random tree, r for the singleton subtree containing the root of $\mathcal{T}$ (recall that the extra permanently infected vertex is not considered part of the infection set) and $\mathbb{E}^{\lambda,\text{slow}}$ for the measure of the slowed SCP with infection rate λ. By Kac's

formula [Norris, 1998, Theorem 3.5.3] applied to the slowed SCP we have for the recovery time R

$$\mathbb{E}_r^{\lambda,\text{slow}}\left(R\middle|\mathcal{T}\right) = \mathbb{E}_\emptyset^{\lambda,\text{slow}}\left(R\middle|\mathcal{T}\right) - \frac{1}{\lambda} = \frac{1}{\lambda\pi_{\lambda\rho}\left(\emptyset\right)} - \frac{1}{\lambda} = \frac{Z_{\lambda\rho}-1}{\lambda},$$

because there is no slowing in the empty tree state. We recall Lemma 7.2.1 and average over $\mathcal{T}$ to see

$$\mathbb{E}_r^{\lambda,\text{slow}}\left(R\right) \leq \rho + \beta\lambda\rho^2 + \frac{e^3}{3\sqrt{6\pi}} \cdot \frac{\lambda^2\rho^3\beta^2}{1-\lambda\rho\beta e}.$$

For this slowed process, if we are at some infection set size $|T| = k \geq 2$, we only return to $k-1$ total infective vertices at rate bounded by $(k-1)\rho^{-k}$ because the root cannot recover while it has an infected child.

Further, if $\rho \in \left(\frac{5}{4}, \frac{4}{3}\right)$ this recovery rate has

$$\max_{k\geq 2}(k-1)\rho^{-k} = \frac{4}{\rho^5}$$

and any sites we infect must recover one-by-one at rate bounded by this maximal rate. Hence

$$\mathbb{E}_r^{\lambda,\text{slow}}\left(R\middle||I_\infty^\emptyset|\right) \geq \rho + \frac{\rho^5}{4}\left(|I_\infty^\emptyset| - 1\right) \tag{7.4}$$

so by taking expectations

$$\rho + \frac{\rho^5}{4}\left(\mathbb{E}_r^{\lambda,\text{slow}}\left(|I_\infty^\emptyset|\right) - 1\right) \leq \mathbb{E}_r^{\lambda,\text{slow}}\left(R\right) \leq \rho + \beta\lambda\rho^2 + \frac{e^3}{3\sqrt{6\pi}} \cdot \frac{\lambda^2\rho^3\beta^2}{1-\lambda\rho\beta e}$$

which we rearrange to obtain

$$\begin{aligned}\frac{\rho^5}{4}\left(\mathbb{E}_r^{\lambda,\text{slow}}\left(|I_\infty^\emptyset|\right) - 1\right) &\leq \rho^2\left(\beta\lambda + \frac{e^3}{3\sqrt{6\pi}} \cdot \frac{\lambda^2\rho\beta^2}{1-\lambda\rho\beta e}\right) \\ \implies \mathbb{E}_r^{\lambda,\text{slow}}\left(|I_\infty^\emptyset|\right) &\leq 1 + \frac{4\beta\lambda}{\rho^3}\left(1 + \frac{e^2}{3\sqrt{6\pi}} \cdot \frac{\lambda\rho\beta e}{1-\lambda\rho\beta e}\right).\end{aligned}$$

Finally we take $\rho \uparrow \frac{4}{3}$ which obtains the claimed result, because we can couple the slowed and unslowed rate-λ SCPs such that they follow the same paths to recovery and so have the same infection set. □

From this bound on the infection spread within a tree, we must bound the number of these infected vertices which update while infected and thus spread the infection to another tree. If the expectation of the number of these vertices is less than 1, by seeing the process of tree sets as its own Galton-Watson process we can argue that the infection

process dies out on the evolving forest.

Lemma 7.2.4. *If $\kappa > 0$ and $\beta\lambda < 0.21$ then the CPEF has*

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\sum_{u\in(\mathbb{N}^+)^n}\mathbb{1}_{|I_\infty^u|\neq 0}\right) < \infty$$

i.e. we expect to infect only finitely many other local trees.

Proof. No vertex can leave the root tree and start a new infected tree unless it was a member of $I_\infty^\emptyset$. Further, if $|I_\infty^\emptyset| - 1$ vertices have updated then the last cannot update, as it would not have any infected neighbours remaining.

Hence we can conclude that this Galton-Watson process of trees is subcritical whenever

$$\mathbb{E}\left(|I_\infty^\emptyset| - 1\right) < 1.$$

By Lemma 7.2.3, we know this is satisfied whenever $\beta\lambda < \frac{3}{4e} > 0.27$ and

$$\frac{27\beta\lambda}{16}\left(1 + \sqrt{\frac{2}{3\pi}}\cdot\frac{2e^3\beta\lambda}{9 - 12\beta\lambda e}\right) < 1$$

which is a quadratic inequality in $\beta\lambda$

$$\begin{aligned}&\frac{243}{16}\beta\lambda - \frac{81e}{4}\beta^2\lambda^2 + \sqrt{\frac{2}{3\pi}}\cdot\frac{27e^3}{8}\beta^2\lambda^2 < 9 - 12\beta\lambda e\\ \iff &\left(\frac{81e}{4} - \sqrt{\frac{2}{3\pi}}\cdot\frac{27e^3}{8}\right)\beta^2\lambda^2 - \left(12e + \frac{243}{16}\right)\beta\lambda + 9 > 0.\end{aligned}$$

We solve the quadratic to find a region $\{\beta\lambda < L\}$, where the limit L is

$$L = \frac{81 + 64e - \sqrt{1152e^3\sqrt{\frac{6}{\pi}} + 4096e^2 - 10368e + 6561}}{12e\left(18 - e^2\sqrt{\frac{6}{\pi}}\right)} > 0.21$$

and hence $\{\beta\lambda < 0.21\}$ is sufficient for subcriticality. □

Having found a region of subcriticality on the CPEF, it remains to check that we can couple this back to the original network problem. The first ingredient is controlling how closely we can approximate the degrees as Poisson.

Lemma 7.2.5 (Theorem 1 of [Barbour and Hall, 1984])**.**

$$d_{TV}\left(\text{Bin}\left(N, \frac{\beta}{N}\right), \text{Pois}(\beta)\right) \leq \frac{\beta\left(1-e^{-\beta}\right)}{N} \leq \frac{\beta}{N}$$

and hence we can couple the two distributions so that they are identical with probability $1-\beta/N$*.*

Proof of Theorem 2.2.8. First fix some large constant $B \in \mathbb{N}^+$ and observe that because the meta-Galton-Watson process in the CPEF is subcritical we expect to see only finitely many tree sets T^u. Hence by Markov's inequality it will see fewer than B tree sets with high probability in B.

Now from Lemma 7.2.3 the first B simulated local trees have an expected total infection size bounded by

$$\left(1+\frac{27\beta\lambda}{16}\left(1+\sqrt{\frac{2}{3\pi}}\cdot\frac{2e^3\beta\lambda}{9-12\beta\lambda e}\right)\right)B < 2B$$

and so with high probability the total number of vertices that are ever infected is bounded by B^2, say, with high probability in B.

It remains to couple the CPEF to the process on the network. In the CPEF trees we have offspring distribution $\text{Pois}(\beta)$ but in the real network these are binomial. Using the bound in Lemma 7.2.5, with high probability in N we can generate B^2 offspring counts from one of these distributions without losing a coupling to the other.

To obtain the precise local distribution of the Erdős-Rényi network we have to give each vertex a uniform label from $[N]$ and *thin* the tree as it is explored by the infection (as in Section 4.3.1, using the order in which the vertices were infected as a hierarchy). We expect in a set of size bounded by B^2 to see a number of label clashes bounded by

$$\binom{B^2}{2}\frac{1}{N} = \Theta\left(\frac{1}{N}\right)$$

and so by Markov's inequality again we see only unique labels, with high probability in N. Hence we have coupled the two infection sets and we know $|I_\infty| \leq B^2$ with high probability in B. Thus for any $\epsilon > 0$ we can find an $M \in \mathbb{N}^+$ such that

$$\sup_{N \geq M} \mathbb{P}\left(|I_\infty| \geq B^2\right) \leq \epsilon$$

and so the set $[1, B^2]$ can contain $1-\epsilon$ mass for a tail of the sequence $(|I_\infty|)_{N\in\mathbb{N}^+}$. □

7.2.2 Supercriticality

For the survival of the epidemic we have more flexibility in choosing a lower bound. In this section we outline two basic SIR couplings that show linear epidemics for large enough infection rate parameter λ. On the network, [Janson et al., 2014] give a complete SIR analysis; here we consider the SIR on a tree and so can simply check the expected number of infected children exceeds 1 and conclude with a more primitive Galton-Watson understanding.

Lemma 7.2.6. *If*

$$\lambda\beta > 1 + 2\kappa + \lambda$$

then the CPEF defined in Definition 4.3.5 has

$$\mathbb{P}(|I_\infty^\emptyset| = \infty) > 0.$$

Proof. We will contain an SIR infection within the SIS on the dynamic graph, in which every vertex is treated as permanently recovered on its first update *attempt* or nontrivial recovery attempt (i.e. its first recovery attempt after becoming infected). Thus the infection is spread down an edge at rate λ and each infected vertex permanently recovers at rate $1 + \kappa$.

Furthermore, each child at the top of an edge in the tree $T^\emptyset$ has independently probability

$$\frac{\kappa}{\kappa + \lambda}$$

to detach itself before it can be infected by its parent. This is true even over multiple infectious periods of the parent: whenever this edge is ready to receive the infection, the child is also ready to update away so the situation is precisely that of competing exponential clocks. We can thus percolate the tree so that the effective offspring distribution is Pois $\left(\frac{\beta\lambda}{\kappa+\lambda}\right)$.

Conditionally on the edge not being "dropped" in this way, the target is infected at rate $\lambda + \kappa$. This SIR model with infection rate $\lambda + \kappa$ and recovery rate $1 + \kappa$ on a tree then infects each child before permanent recovery with probability

$$\frac{\lambda + \kappa}{1 + \kappa + \lambda + \kappa} = \frac{\lambda + \kappa}{1 + \lambda + 2\kappa}$$

by again considering two competing exponential clocks. On the Pois $\left(\frac{\beta\lambda}{\kappa+\lambda}\right)$-Galton-

Watson tree, then, an infected vertex expects to infect at least 1 child if

$$\frac{\lambda+\kappa}{1+\lambda+2\kappa}\cdot\frac{\beta\lambda}{\lambda+\kappa}=\frac{\beta\lambda}{1+2\kappa+\lambda}>1$$

which is true by assumption. Thus we contain a supercritical Galton-Watson process of infected vertices within successive generations of the underlying Galton-Watson tree, and so there is positive probability that the process will infect infinitely many vertices. □

This survival on the root tree of the CPEF is the heart of the proof, but again we must translate it back to a claim about the finite network to prove Theorem 2.2.10.

Proof of Theorem 2.2.10. As in Section 4.3.1, we explore the network with the infection, keeping a set of revealed vertices $R_t \subset [N]$ such that each vertex is revealed on infection and unrevealed on updating while healthy.

Because the SIR lower bound (as in the previous lemma) only considers infected vertices before they check their neighbourhood for infected vertices, we can consider it on a subtree of the full revealed graph – survival on this subtree will force an epidemic in the more complicated infection that contains it.

Recall that the distribution of offspring in the exploration of Section 4.3.1 was

$$\mathrm{Bin}\left(N-|R_t|,\frac{\beta}{N}\right)$$

and thus while $|R_t|<\epsilon N$ we expect at least $\beta(1-\epsilon)$ unrevealed offspring from every vertex exploration.

To further incorporate the dropping of unrevealed offspring as in the previous lemma, we must keep children unique. Keep a *claimed* set $C_t \subseteq [N]$ of vertices which were ever adjacent to infected vertices, note that $R_t \subseteq C_t$. While $|C_t|<\epsilon N$, we expect at least $\beta(1-\epsilon)$ unrevealed offspring *that won't be claimed by another vertex* from every vertex exploration.

Hence if we define

$$\tau := \inf_{t>0}\{C_t \geq \epsilon N\},$$

we can contain an SIR infection in our graph up to the time τ, and if further

$$\frac{\beta\lambda(1-\epsilon)}{1+2\kappa+\lambda}>1$$

then this infection has an asymptotically positive probability to survive, which we make true by setting ϵ sufficiently small. Clearly, the SIR infection cannot survive to infect

infinitely many distinct vertices in $[N]$ and so we conclude

$$\liminf_{N\to\infty} \mathbb{P}_1 (\tau < \infty) > 0.$$

If we do find $\tau < \infty$ then we have some set $|C_\tau| \geq \epsilon N$ where every $v \in C_\tau$ was exposed at time $t_v \leq \tau$ to some $w_v \in I_{t_v}$. Every such exposure has an event $E(v)$ with probability $\frac{\lambda}{1+\lambda+2\kappa}$ to infect v before the next recovery time of w_v or a possible update time of either vertex.

Because an unclaimed vertex cannot update, v can only become claimed by either an infection or an update at w_v. Hence for $v_1, v_2 \in C_\tau$ we either have $E(v_1), E(v_2)$ independent, or they were claimed by the same vertex at the same time. Because we claim at most

$$\text{Bin}\left(N, \frac{\beta}{N}\right) \preceq \text{Pois}(\beta + \epsilon)$$

at a time (see e.g. [Klenke and Mattner, 2010, Theorem 1.1(f)]) we can conclude that with high probability there are at least

$$\frac{\epsilon N}{\beta + 2\epsilon}$$

vertices $v \in C_\tau$ with mutually independent exposure events $E(v)$, because otherwise this bound of the Poisson sum would be displaying a large deviation. By binomial concentration, increasing the constant 2ϵ to 3ϵ, with high probability at least

$$\frac{\epsilon N}{\beta + 3\epsilon} \cdot \frac{\lambda}{1 + \lambda + 2\kappa}$$

of these events will then occur and thus on the event $\{\tau < \infty\}$ we find the epidemic event $E_{\frac{\epsilon}{\beta+3\epsilon}\frac{\lambda}{1+\lambda+2\kappa}}$ with high probability in N. Thus we have found positive probability of an epidemic event, whether by the previous SIR lower bound or otherwise simply by observing a large level of exposure that the lower bound did not control. □

Appendix A

Counting Rooted Subtrees

The following result is proven by analytic combinatorial methods on Mathematics Stack Exchange [Riedel,]. We reproduce the proof here for permanence and to provide extra explanation.

Proposition A.0.1. *Take $d \geq 3$. In $\mathcal{U}$ the set of rooted subtrees of the rooted d-regular tree T_d, we have*

$$\frac{d}{k-1}\binom{k(d-1)}{k-2} =: C_k$$

rooted subtrees of size k.

Proof. We consider two combinatorial classes: $\mathcal{T}$ containing the rooted subtrees of the infinite rooted $(d-1)$-ary tree, in which every vertex has $d-1$ children and so the root has degree $d-1$; and $\mathcal{U}$ containing the the rooted subtrees of the infinite rooted d-regular tree in which the root has d children.

We describe the class $\mathcal{T}$ via the recursive equation

$$\mathcal{T} = \mathcal{E} + \mathcal{Z} \times \mathcal{T}^{d-1}$$

where $\mathcal{E}$ represents the empty tree, and $\mathcal{Z}$ denotes a vertex in the root, which could then have any $d-1$ other trees from $\mathcal{T}$ attached to the $d-1$ child branches. Then $\mathcal{U}$ has a similar description reflecting that only at the root were we permitted an extra choice

$$\mathcal{U} = \mathcal{E} + \mathcal{Z} \times \mathcal{T}^{d}.$$

These recursive equations have associated functional equations for the generating

functions $T(z)$, $U(z)$ of the respective classes [Flajolet and Sedgewick, 2009]

$$T(z) = 1 + zT(z)^{d-1}, \qquad U(z) = 1 + zT(z)^d,$$

from which by substitution we find

$$U(z) = T(z)^2 - T(z) + 1.$$

We are interested, for $k \geq 1$, in the coefficient C_k of z^k in $U(z)$ which we extract by Cauchy's integral formula as in [Flajolet and Sedgewick, 2009, Appendix B2] for a small simple positive loop Γ encircling 0

$$\begin{aligned}[z^k]U(z) = \frac{1}{k}[z^{k-1}]U'(z) &= \frac{1}{k}[z^{k-1}]T'(z)(2T(z)-1) \\ &= \frac{1}{2\pi ki}\oint_\Gamma \frac{2T(z)-1}{z^k}T'(z)\mathrm{d}z \\ &= \frac{1}{2\pi ki}\oint_{1+\Gamma} \frac{w^{k(d-1)}(2w-1)}{(w-1)^k}\mathrm{d}w\end{aligned}$$

by substituting $w = T(z)$ and using (because there is one empty tree and one of size 1) that $T(z) = 1 + z + O\left(z^2\right)$ to deform a sufficiently small loop. Translate this integral back to the origin to see

$$\begin{aligned}C_k &= \frac{1}{2\pi ki}\oint_\Gamma \frac{(w+1)^{k(d-1)}(2w+1)}{w^k}\mathrm{d}w = \frac{1}{k}[w^{k-1}](w+1)^{k(d-1)}(2w+1) \\ &= \frac{1}{k}\left(2\binom{k(d-1)}{k-2} + \binom{k(d-1)}{k-1}\right) \\ &= \frac{1}{k}\left(2\binom{k(d-1)}{k-2} + \binom{k(d-1)}{k-2}\frac{k(d-1)-(k-2)}{k-1}\right) \\ &= \frac{d}{k-1}\binom{k(d-1)}{k-2}\end{aligned}$$

which is the claimed expression. □

Bibliography

[Abdullah et al., 2012] Abdullah, M., Cooper, C., and Frieze, A. (2012). Cover time of a random graph with given degree sequence. *Discrete Mathematics*, 312(21):3146–3163.

[Aldous, 1989] Aldous, D. (1989). Probability approximations via the poisson clumping heuristic. *Applied mathematical sciences*, 77.

[Aldous and Fill, 2002] Aldous, D. and Fill, J. A. (2002). Reversible markov chains and random walks on graphs. Unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/$\sim$aldous/RWG/book.html`.

[Alon and Spencer, 2016] Alon, N. and Spencer, J. H. (2016). *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition.

[Athreya, 2012] Athreya, K. (2012). Coalescence in the recent past in rapidly growing populations. *Stochastic Processes and their Applications*, 122(11):3757–3766.

[Bagley, 1982] Bagley, J. (1982). Asymptotic properties of subcritical galton-watson processes. *Journal of Applied Probability*, 19(3):510–517.

[Bailey, 1975] Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

[Ball et al., 2019] Ball, F., Britton, T., Leung, K. Y., and Sirl, D. (2019). A stochastic sir network epidemic model with preventive dropping of edges. *Journal of mathematical biology*, 78(6):1875–1951.

[Barabási and Albert, 1995] Barabási, A.-L. and Albert, R. (1995). Emergence of scaling in random networks. *Phys. Rev. Lett*, 74:3273.

[Barabási et al., 2006] Barabási, A.-L., Newman, M., and Watts, D. J., editors (2006). *The structure and dynamics of networks*. Princeton Studies in Complexity. Princeton University Press, Princeton, NJ.

[Barbour and Hall, 1984] Barbour, A. D. and Hall, P. (1984). On the rate of poisson convergence. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 95, pages 473–480. Cambridge University Press.

[Benjamini et al., 2014] Benjamini, I., Kozma, G., and Wormald, N. (2014). The mixing time of the giant component of a random graph. *Random Structures & Algorithms*, 45(3):383–407.

[Berestycki et al., 2018] Berestycki, N., Lubetzky, E., Peres, Y., Sly, A., et al. (2018). Random walks on the random graph. *The Annals of Probability*, 46(1):456–490.

[Bhamidi et al., 2019] Bhamidi, S., Nam, D., Nguyen, O., and Sly, A. (2019). Survival and extinction of epidemics on random graphs with general degrees. *arXiv preprint arXiv:1902.03263*.

[Bingham and Doney, 1974] Bingham, N. H. and Doney, R. A. (1974). Asymptotic properties of supercritical branching processes i: The galton-watson process. *Advances in Applied Probability*, 6(4):711–731.

[Bollobás et al., 2007] Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122.

[Bollobás and Riordan, 2015] Bollobás, B. and Riordan, O. (2015). An old approach to the giant component problem. *Journal of Combinatorial Theory, Series B*, 113:236–260.

[Chatterjee et al., 2009] Chatterjee, S., Durrett, R., et al. (2009). Contact processes on random graphs with power law degree distributions have critical value 0. *The Annals of Probability*, 37(6):2332–2356.

[Chen et al., 2016] Chen, Y.-T., Choi, J., and Cox, J. T. (2016). On the convergence of densities of finite voter models to the wright–fisher diffusion. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 286–322. Institut Henri Poincaré.

[Chin et al., 2018] Chin, A. J., Gordon, G., MacPhee, K. J., and Vincent, C. (2018). Subtrees of graphs. *Journal of Graph Theory*, 89(4):413–438.

[Chung and Lu, 2006] Chung, F. and Lu, L. (2006). *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI.

[Clifford and Sudbury, 1973] Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, 60(3):581–588.

[Cooper et al., 2016] Cooper, C., Dyer, M., Frieze, A., and Rivera, N. (2016). Discordant voting processes on finite graphs. *arXiv preprint arXiv:1604.06884*.

[Cooper and Rivera, 2016] Cooper, C. and Rivera, N. (2016). The linear voting model. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 144, 12. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern.

[Davies, 1978] Davies, P. (1978). The simple branching process: a note on convergence when the mean is infinite. *Journal of Applied Probability*, pages 466–480.

[Ding et al., 2014] Ding, J., Lubetzky, E., and Peres, Y. (2014). Anatomy of the giant component: The strictly supercritical regime. *European Journal of Combinatorics*, 35:155–168.

[Durrett, 2010] Durrett, R. (2010). Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences*, 107(10):4491–4498.

[Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

[Erdős and Rényi, 1959] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen*, 6:290–297.

[Erdős and Rényi, 1960] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61.

[Erdős and Rényi, 1961] Erdős, P. and Rényi, A. (1961). On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267.

[Etheridge, 2011] Etheridge, A. (2011). *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[Fernholz and Ramachandran, 2007] Fernholz, D. and Ramachandran, V. (2007). The diameter of sparse random graphs. *Random Structures & Algorithms*, 31(4):482–516.

[Fernley and Ortgiese, 2019] Fernley, J. and Ortgiese, M. (2019). Voter models on subcritical inhomogeneous random graphs. *arXiv preprint arXiv:1911.13187*.

[Fill, 1991] Fill, J. A. (1991). Time to stationarity for a continuous-time markov chain. *Probability in the Engineering and Informational Sciences*, 5(1):61–76.

[Flajolet and Sedgewick, 2009] Flajolet, P. and Sedgewick, R. (2009). *Analytic combinatorics*. cambridge University press.

[Fountoulakis and Reed, 2008] Fountoulakis, N. and Reed, B. A. (2008). The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures & Algorithms*, 33(1):68–86.

[Gilbert, 1959] Gilbert, E. N. (1959). Random graphs. *Ann. Math. Statist.*, 30:1141–1144.

[Gut, 2013] Gut, A. (2013). *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, second edition.

[Harris, 1960] Harris, T. E. (1960). A lower bound for the critical probability in a certain percolation process. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge University Press.

[Harris, 1963] Harris, T. E. (1963). *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J.

[Harris, 1974] Harris, T. E. (1974). Contact interactions on a lattice. *Ann. Probability*, 2:969–988.

[Hoeffding, 1963] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

[Holley and Liggett, 1975] Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663.

[Huang and Durrett, 2018] Huang, X. and Durrett, R. (2018). The contact process on random graphs and galton-watson trees. *arXiv preprint arXiv:1810.06040*.

[Jacob et al., 2019] Jacob, E., Linker, A., and Mörters, P. (2019). Metastability of the contact process on fast evolving scale-free networks. *Ann. Appl. Probab.*, 29(5):2654–2699.

[Jacob and Mörters, 2017] Jacob, E. and Mörters, P. (2017). The contact process on scale-free networks evolving by vertex updating. *Royal Society Open Science*, 4(5):170081.

[Jansen et al., 2014] Jansen, S., Kurt, N., et al. (2014). On the notion (s) of duality for markov processes. *Probability surveys*, 11:59–120.

[Janson, 2004] Janson, S. (2004). Functional limit theorems for multitype branching processes and generalized pólya urns. *Stochastic Processes and their Applications*, 110(2):177–245.

[Janson, 2008] Janson, S. (2008). The largest component in a subcritical random graph with a power law degree distribution. *The Annals of Applied Probability*, 18(4):1651–1668.

[Janson, 2012] Janson, S. (2012). Simply generated trees, conditioned galton–watson trees, random allocations and condensation. *Probability Surveys*, 9:103–252.

[Janson et al., 2014] Janson, S., Luczak, M., and Windridge, P. (2014). Law of large numbers for the sir epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 45(4):726–763.

[Jonsson and Stefánsson, 2011] Jonsson, T. and Stefánsson, S. Ö. (2011). Condensation in nongeneric trees. *Journal of Statistical Physics*, 142(2):277–313.

[Kallenberg, 2006] Kallenberg, O. (2006). *Foundations of modern probability*. Springer Science & Business Media.

[Kanade et al., 2016] Kanade, V., Mallmann-Trenn, F., and Sauerwald, T. (2016). On coalescence time in graphs–when is coalescing as fast as meeting? *arXiv preprint arXiv:1611.02460*.

[Kermack and McKendrick, 1927] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.

[Kesten and Stigum, 1966] Kesten, H. and Stigum, B. P. (1966). A limit theorem for multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, 37(5):1211–1223.

[Keynes, 1921] Keynes, J. M. (1921). *A treatise on probability*. Macmillan and Company, limited.

[Klenke and Mattner, 2010] Klenke, A. and Mattner, L. (2010). Stochastic ordering of classical discrete distributions. *Advances in Applied probability*, 42(2):392–410.

[Levin et al., 2009] Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov chains and mixing times*. American Mathematical Society, Providence, RI. With a chapter by James G. Propp and David B. Wilson.

[Liggett, 1985] Liggett, T. M. (1985). *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York.

[Liggett, 1999] Liggett, T. M. (1999). *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

[Mach et al., 2020] Mach, T., Sturm, A., and Swart, J. M. (2020). Recursive tree processes and the mean-field limit of stochastic flows. *Electron. J. Probab.*, 25:Paper No. 61, 63.

[Moinet et al., 2018] Moinet, A., Barrat, A., and Satorras, R. P. (2018). Generalized voter-like model on activity driven networks with attractiveness. *arXiv preprint arXiv:1804.07476*.

[Moran, 1958] Moran, P. A. P. (1958). Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, volume 54, pages 60–71. Cambridge University Press.

[Mountford et al., 2013] Mountford, T., Valesin, D., Yao, Q., et al. (2013). Metastable densities for the contact process on power law random graphs. *Electronic Journal of Probability*, 18.

[Mourrat and Valesin, 2016] Mourrat, J.-C. and Valesin, D. (2016). Phase transition of the contact process on random regular graphs. *Electron. J. Probab.*, 21:Paper No. 31, 17.

[Nam et al., 2019] Nam, D., Nguyen, O., and Sly, A. (2019). Critical value asymptotics for the contact process on random graphs. *arXiv preprint arXiv:1910.13958*.

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

[Norris, 1998] Norris, J. R. (1998). *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. Reprint of 1997 original.

[Norros and Reittu, 2006] Norros, I. and Reittu, H. (2006). On a conditionally poissonian graph process. *Advances in Applied Probability*, 38(1):59–75.

[Oliveira, 2012] Oliveira, R. (2012). On the coalescence time of reversible random walks. *Transactions of the American Mathematical Society*, 364(4):2109–2128.

[Oliveira, 2013] Oliveira, R. I. (2013). Mean field conditions for coalescing random walks. *The Annals of Probability*, 41(5):3420–3461.

[Pakes and Dekking, 1991] Pakes, A. G. and Dekking, F. (1991). On family trees and subtrees of simple branching processes. *Journal of Theoretical Probability*, 4(2):353–369.

[Pastor-Satorras and Vespignani, 2002] Pastor-Satorras, R. and Vespignani, A. (2002). Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):035108.

[Pemantle, 1992] Pemantle, R. (1992). The contact process on trees. *The Annals of Probability*, 20(4):2089–2116.

[Peres et al., 2017] Peres, Y., Sauerwald, T., Sousi, P., Stauffer, A., et al. (2017). Intersection and mixing times for reversible chains. *Electronic Journal of Probability*, 22.

[Riedel,] Riedel, M. Number of rooted subtrees of given size in infinite d-regular tree. Mathematics Stack Exchange. https://math.stackexchange.com/q/855970 (version: 02/23/2020).

[Saloff-Coste, 1997] Saloff-Coste, L. (1997). Lectures on finite markov chains. In *Lectures on probability theory and statistics*, pages 301–413. Springer.

[Sinclair, 1992] Sinclair, A. (1992). Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, probability and Computing*, 1(4):351–370.

[Sood et al., 2008] Sood, V., Antal, T., and Redner, S. (2008). Voter models on heterogeneous networks. *Physical Review E*, 77(4):041121.

[van der Hofstad, 2016] van der Hofstad, R. (2016). *Random graphs and complex networks*, volume 43. Cambridge university press.

[van der Hofstad, 2020] van der Hofstad, R. (2020). *Random graphs and complex networks: Volume II*.