Random Graphs With Correlation Structure

David Binnie Penman

This thesis is submitted in partial fulfillment of the requirements for the Ph.D degree in the Probability and Statistics Section, School of Mathematics, University of Sheffield in December 1997.

Accepted in February 1998.

PAGE NUMBERING AS ORIGINAL

Summary

Summary of: Random Graphs with Correlation Structure

Author; David Binnie Penman

Thesis submitted in partial fulfillment of the requirements for the Ph.D degree in the Probability and Statistics Section, School of Mathematics, University of Sheffield in December 1997.

In this thesis we consider models of random graphs where, unlike in the classical models $G(n,p)$ the probability of an edge arising can be correlated with that of other edges arising. Attention focuses on graphs whose vertices are each assigned a colour (type) at random and where edges between differently coloured vertices subsequently arise with different probabilities (so-called **RRC graphs**), especially the special case with two colours. Various properties of these graphs are considered, often by comparing and contrasting them with the classical model with the same probability of each particular edge existing. Topics examined include the probabilities of trees and cycles, how the joint probability of two subgraphs compares with the product of their probabilities, the number of edges in the graph (including large deviations results), connectedness, connectivity, the number and order of complete graphs and cliques, and tournaments with correlation structure.

Contents

1 Introduction

1.1 Summary

In this thesis we consider models of random graphs where, unlike in the classical models $G(n,p)$, the probability of an edge arising can be correlated with that of other edges arising. Attention focuses on graphs whose vertices are each assigned a colour (type) at random and where edges between differently coloured vertices subsequently arise with different probabilities (so-called **RRC graphs**), especially the special case with two colours. Various properties of these graphs are considered, often by comparing and contrasting them with the classical model with the same probability of each particular edge existing. Topics examined include the probabilities of trees and cycles, how the joint probability of two subgraphs compares with the product of their probabilities, the number of edges in the graph (including large deviations results), connectedness, connectivity, the number and order of complete graphs and cliques, and tournaments with correlation structure.

1.2 Declaration

This thesis is submitted in partial fulfilment of the requirements for the Ph.D degree of the University of Sheffield. No part of this thesis has been submitted before to any university for any degree, diploma or other award.

1.3 Acknowledgements

The greatest debt is, of course, to my two supervisors, John Biggins and Chris Cannings for their many helpful suggestions, their detailed criticisms of various drafts of the material, and their constant encouragement. I am also grateful to the University of Sheffield for its financial support through a two year University Studentship. I am also grateful to all my family, and especially my parents, for their support in numerous ways.

1.4 The basic notions in random graphs

This thesis aims to initiate the study of some models of random graphs with a correlation structure. A **graph** $G = (V, E)$ is a finite set $V = V(G)$ of **vertices** and a set E of **edges** between certain of these vertices; two vertices with an edge between them are said to be adjacent. We do not allow edges from a vertex to itself, so that there are no loops, and normally forbid **multigraphs** where there can be multiple edges between two vertices. Occasionally the edges will have an orientation, that is the edge between i and j is directed from $i \to j$ or $j \to i$; this is a **digraph**.

One mechanism for generating **random graphs** occurs where the set E is chosen by some stochastic mechanism, the set V having been fixed in advance. (Randomness in the order of V is also possible, though less studied). It is technically convenient (for counting arguments) to study **labelled graphs** on $V = \{1, 2, ...n\}$, that is to distinguish between isomorphic graphs which are not the same labelled graph. Then a model of random graphs is a rule assigning probabilities to each of the $2^{n(n-1)/2}$ possible labelled graphs on n vertices. There are obviously uncountably many ways in which one could do this, and very little can be said in such generality. To make progress one needs an amenable probability distribution; and it is clearly desirable that interesting events should be independent, to facilitate calculations of probabilities. Since the basic relation in the theory is adjacency, we might guess the interesting events are whether or not each possible edge occurs; these remarks show why much previous work on random graphs has been on the model $G(n, p(n))$ (or $G_{p(n)}$ if n is clear; we shall describe this model as the **classical model**), where each possible edge arises with probability $p(n)$ independently of other edges. (Often this is written $G(n, p)$ taking the dependence of p on n as read; we shall try to avoid this). Another common model is $G(n, M(n))$, where $M(n)$ of the $n(n-1)/2$ possible edges occur, all sets of $M(n)$ edges being equally likely; the main reason why this model is rather harder to study is that it has less good independence properties. These models are studied extensively in the standard text [B] where many results on the subject, whose study was initiated by Erdos and Renyi, are presented in a unified way; we shall often refer to this book.

There are (at least) two main strands in random graph theory. The first is proving the existence of graphs with some property $\wp$, by showing that, in some model, $P\{G \text{ has } \wp\} > 0$. A standard example is graphs with arbitrarily high girth and chromatic number (this is noteworthy as a graph of large girth

looks locally like a tree, and trees have chromatic number 2). Arguments of this type can in principle be replaced by counting arguments, but in practice this is often impractical, so the probabilistic method yields new insight. In fact it is often very hard to get graphs which are explicit examples, even if the random graph argument indicates there are many such graphs; for example, concerning graphs of high girth and chromatic number, no explicit examples were known for many years after the probabilistic proof; the first construction by Lovasz required ideas from multigraph theory, and ideas from other parts of mathematics (the Weil estimates from number theory) were needed to get the first reasonably intuitive examples (so-called Ramanujan graphs).

The second strand, studied more in [B], is **asymptotic** behaviour when the number of vertices n is large; as exact enumerative formulae are usually intractable, we instead consider some family of probability spaces such as $\{G(n,p(n))\}$ for all n, or $\{G(n,M(n))\}$, and examine the proportion of graphs in $G(n,p(n))$ having $\wp$ as $n \to \infty$, by suitable approximations. We say **almost every** (a.e.) graph in $G(n,p)$ has $\wp$ if $\lim_{n\to\infty} P\{G(n,p) \text{ has } \wp\} = 1$. We note that this terminology, though standard in random graphs, differs from the use of a.e. in general probability theory; it is really a convergence in probability notion.

Of great importance is the discovery that many interesting graph properties (but not all; see for example [T] and, to emphasise that the property **is** of interest, [CO]) have a $0-1$ law in the following loose sense; given $p(n)$ either a.e. graph in $G(n,p(n))$ has $\wp$ or a.e. graph does not have $\wp$. Often there is a **threshold function**, that is a function $p^*(n)$ such that

$$\lim_{n\to\infty} \frac{p(n)}{p^*(n)} = \infty \Rightarrow \lim_{n\to\infty} P\{G_p \text{ has } \wp\} = 1,$$

$$\text{but } \lim_{n\to\infty} \frac{p(n)}{p^*(n)} = 0 \Rightarrow \lim_{n\to\infty} P\{G_p \in \wp\} = 0$$

(p^*(n) is not unique but this matters little). Indeed Bollobas and Thomason show every **monotone** property has a threshold ($\wp$ is monotone if and only if, given subgraphs H and K of G, we have

$$(K \leq H \leq G) \text{ and } (K \text{ has } \wp) \Rightarrow (H \text{ has } \wp)\,;$$

if the property $\wp$ is monotone then $P\{G_p \text{ has } \wp\}$ is an increasing function of p).

1.5 A new model; our approach to its study

The assumption that the edges arise independently and equiprobably is clearly desirable to facilitate probability calculus. However, quite apart from the obvious purely mathematical desire to understand what happens in more general circumstances, this assumption is clearly problematic in practice; for example, we may know that the vertices are of different types and that the probability an edge arises depends on the types of its two vertices. Thus we want to consider models, generalising $G(n, p(n))$, where whether an edge arises depends on the types of its two vertices. We chose to study such models because, while they are easily seen to differ from G_p in many ways, they retain a lot of independence to make calculations work, since conditional on the types of the vertices, edges still arise independently, though not now usually equiprobably.

The types are conveniently represented by assigning one of k different colours to the vertices, provided it is clearly understood that these colourings have nothing to do with **proper colourings** where we assign a colour to each vertex so that no adjacent vertices are the same colour. We also emphasise that the objects of study are graphs, not coloured graphs; theorems will discuss the probability that a graph has some property, not that a graph with some kind of colouring has it (although in proofs we will of course consider the colours). Sloganising, we see things in monochrome.

A further question is whether one wants to specify the number of vertices of each colour randomly or deterministically. There are arguments both ways, but generally models where colours are assigned to each vertex independently (in the cases we shall consider, from some multinomial distribution) have better exchangeability properties, in that the probability that an edge arises is the same for all edges, allowing us to compare the behaviour of graph invariants in our model and a classical model with that probability, which will be denoted by the good classical letter α. We shall call such graphs **RRC graphs**, standing for **random randomly coloured graphs**, to reflect the two levels at which randomness is present, and we shall talk about RRC or new models, as opposed to the classical ones. In fact, we shall often be comparing the behaviour of a whole range of new models, as the probabilities of edges between vertices of given types vary, subject to the constraint that α stays fixed, with a classical model with probability α. Of course, sometimes we can only prove results for certain types of RRC models, for example, those with just two colours.

1.6 What is already known about this subject?

On the subject of classical random graphs, a great deal is known; several important developments have occurred since [B] appeared in the mid-1980s; [B1] summarises some of these, but the subject is still evolving.

On the other hand, there seems, on the basis of a literature search and a personal communication from A Thomason, to be no literature on RRC graphs. There are scattered papers on the situation where the edge between i and j arises with probability p_{ij}, independently of all other edges (the model denoted $G\{n,(p_{ij})\}$ in [B], II.1); see [Ke], [Ko] and [Ju]. Thus, if we condition on the colouring of the vertices, we can (and sometimes will) use their results.

Models closely related to some of ours have been studied in [AK], which is basically concerned with the behaviour (for large n) of colouring algorithms for graphs which consist of k (k constant) blocks of nearly equal size, no edges between vertices in the same block and edges between blocks arising independently with probability $p = c/n$ for some constant c. However the questions considered there are algorithmic and do not directly bear on this material.

There has been some study of other models where edges are no longer independent; for example, the so called random cluster model studied in [BGJ], where, with n vertices, the probability that the set E of edges arises is (up to a normalising constant to make the expression a probability function) $p^{|E|}(1-p)^{n(n-1)/2-|E|}q^{c(V,E)}$ where $c(V,E)$ is the number of **components** of the graph $G = (V,E)$. This model is motivated by ideas from statistical physics. Another model, where one generates n independent random points uniformly on $[0,1]^d$ and says two points are adjacent if and only if their l_∞ distance is at most some prescribed value $x(n) \in [0,1]$, has recently been studied by Appel and Russo (see [AR] and references therein), who discuss the rates of convergence or divergence, as $n \to \infty$ and $x(n)$ varies with n, of the maximum and minimum degrees and the connectivity; other recent work on this model has been done by Penrose and others. It is easy to see that these models have a quite different correlation structure from ours.

Of course, there has also been study of models such as all r-regular graphs, ([B], II.4), which in some sense have a correlation structure; however we shall only consider models which assign a probability to every graph on n vertices (except in Chapter 8). We mention also recent work by Ruczinski, Wormald and others on graph processes with bounded maximum degree (see [RW] and references therein); again the correlation structure there is very different from

ours.

In addition to the fact that there is a lack of previous literature on such models, it is true that many probabilities (for example, the probability of a cycle) which are trivial to work out classically, are less obvious in our models. Hence some chapters consider such preliminary problems, and so have a rather different flavour from much of the random graphs literature, though later chapters contain results which look more like the usual theory. Again due to the lack of previous knowledge, we sometimes get a feel for the behaviour of invariants by simple simulation experiments (usually based on FORTRAN programs); however, as the aim is always to get analytic results, these are not critical, so we do not discuss them in detail. Also we sometimes use MAPLE to simplify cumbersome algebraic expressions, mostly in counterexamples to putative theorems rather than being critical to the development. We have marked such equations by using some phrase such as 'using computer simplification'; the technique is probably no more unreliable than hand calculation, though we have usually tried to check some simple cases of such formulae by hand. Similarily, some matrix computations have been performed in MATLAB.

We mention one way in which our models will be harder to study than the classical one. If $\wp$ is a property of graphs with threshold probability $\alpha^*(n)$ in G_α, and we have an RRC model with probabilities $p_{ij}(n)$ then, setting $r(n) = \min_{1\leq i,j\leq k}\{p_{ij}(n)\}$ and $t(n) = \max_{1\leq i,j\leq k}\{p_{ij}\}$, it is clear that

$$\lim_{n\to\infty}\frac{r(n)}{\alpha^*(n)} = \infty \Rightarrow \lim_{n\to\infty} P\{G_\alpha \text{ has } \wp\} = 1 \text{ but}$$

$$\lim_{n\to\infty}\frac{t(n)}{\alpha^*(n)} = 0 \Rightarrow \lim_{n\to\infty} P\{G \text{ has } \wp\} = 0.$$

However in our situation, some of the $p_{ij}(n)$ will be larger than $\alpha^*(n)$ and others will be smaller, so the behaviour is less easy to predict.

1.7 Contents of the various chapters of this thesis

We now outline the areas examined in this thesis, discussing the main results proved; these are all original results (so far as we know), though of course some of them will depend for the proofs on the work of other authors. In the first four chapters, we consider how the correlation structure is and is not manifested in our models.

In Chapter 2 we consider the probabilities of simple configurations of edges, such as trees and cycles, arising, and see whether they are the same as, greater than, or less than classically. The main original results are a classification of when trees have the same probability of arising as classically, several partial results towards the conjecture that the probability that any tree arises is always at least as large as classically, formulae for the probability of a cycle, and some consequences including the fact that the probabilities of cycles distinguish between classical and new models except in trivial cases, fairly precise results on when cycles are more or less likely than classically, and an asymptotic estimate of the probability of a cycle.

This leads naturally to considering in Chapter 3 how the joint probability of two subgraphs arising compares with the product of their probabilities of arising. The main result is Theorem 3.1, stating that in a certain model $G_{p,q}$ defined in section 1.8 below, provided $p > q$, the joint probability is always at least as large as the product of the probabilities. Much of the rest of the chapter consists of complements to and extensions of that result; what happens if $q > p$, a (partial) extension to some more general models, and proving that certain putative extensions are not possible. We also discuss the role of the FKG and Janson inequalities, which are used to study such questions in the classical model; it is shown that these inequalities do not hold in general in our setup, but that some at least of their consequences can be recovered by different methods.

In Chapter 4 we forget about the pattern of the edges and concentrate simply on their number, mainly by relating its moments in the new and old models; results include comparison of the moments, especially the mean and variance for all models, a stochastic dominance result for one class of models, a discussion of properties generalising independence for the edges leading to a central limit theorem and some discussion of Poisson approximation, and an estimate of the maximum degree in $G_{p,q}$ by exploiting results of Bollobas on the classical model.

In Chapter 5, we study the more technical subject of **large deviations** in the number of edges; a form of the Gartner-Ellis theorem is used to obtain a complete result for $G_{p,q}$, despite some technical subtleties, some consequences on the non-existence of martingales are noted, and some partial results for general models are obtained by exploiting a link with the seemingly unrelated area of ESS theory.

In Chapter 6 we study connectedness; G is **connected** if it has just one component. Even if G is disconnected, we are interested in the number of

components and their orders. We also study the various related measures of **connectivity**, that is how fragile the connectedness of a connected graph is. The main results include theorems giving the probability of connectedness in $G_{p,q}$ for small values of p and q, a result (using a result of Juhasz) on one of the various measures of connectivity in $G_{p,q}$ and some consequences for the other measures, and some results on the diameter of our graphs.

In Chapter 7 we study cliques; a **complete subgraph** of G is $A \subset V(G)$ with all possible edges between vertices in A existing in E; a **clique** a complete subgraph contained in no other complete subgraph. Its order is the number of vertices in it. We compare numbers and orders of cliques in our and the classical model, and also discuss related topics such as independent sets and chromatic numbers. The main results are formulae for the expected numbers of these in our models, and some discussion of their asymptotic behaviour, with partial results on the variability of the number of complete graphs, and the fact that the distribution of clique sizes is often multimodal. We also consider the evolving clique, a birth process to generate a complete graph, and obtain some results comparing how it grows in our models with classically.

So far we have discussed only simple graphs, but there are other models where similar ideas could be useful. One such is **tournaments**, that is complete graphs made into digraphs by orienting each edge $i - j$ from i to j or vice versa (not both). There is a classical model of random tournaments, where the edge between i and j goes $i \to j$ or $j \to i$ equiprobably, with the orientations of different edges being independent. Here, by analogy with our models of undirected graphs, we make the orientation of an edge dependent on the random colours of the two vertices involved. We investigate some basic properties of such models in Chapter 8; main results include formulae for the probability of a (directed) cycle and results on when it is more or less likely than classically; comparison of the joint probability of two subgraphs with the product of their probabilities in a certain model; results on the number of (directed) 3-cycles and the probability of irreducibility; and the outdegrees of the vertices, including large deviations.

Finally, in Chapter 9, we summarise where we have got to and give some pointers to future topics meriting investigation. We also make some brief remarks about applications and statistical questions related to the work.

1.8 Basics, especially notation.

The main purpose of this subsection is to introduce some notation which will be used throughout the thesis. Much of it will be fairly standard. The word positive means strictly greater than zero; non-negative means positive or zero, with similar conventions for negative and non-positive. Similarily, functions or sequences will be called increasing if they do so strictly and non-decreasing if they are only increasing in the weak sense. $[a, b]$ denotes the closed interval $a \leq x \leq b$ and (a, b) the open interval $a < x < b$. Vectors will be written in boldface, for example $\mathbf{s}$, and matrices will be denoted by capital letters. Logarithms are to the base e, that is natural logarithms, unless another base a is indicated as a subscript, for example $\log_a(n)$. The rth derivative of $f(x)$ will often be denoted $f^{(r)}(x)$. The symbol $\simeq$ will be used to indicate approximate equality, often where higher order terms of an expansion are of little interest.

There will be no global numbering of equations; however $(*)$, $(\&)$ and similar markers will occasionally be used on equations which will be frequently discussed within that chapter. The symbol $\bullet$ denotes the end of a proof (or absence of one if the result is obvious in its context).

Graphs, as stated before, are finite, without multiple edges or loops, and undirected (many authors call these simple graphs), except edges are directed in Chapter 8. We consider **labelled** graphs on vertex set $V = \{1, 2, ...n\}$ throughout this thesis, unless explicitly stated otherwise; thus, if we talk of the probability that some collection of edges arises, we mean the probability that some particular collection of labelled edges arises, not that some graph isomorphic to that collection arises. As mentioned before, this simplifies counting arguments, but is a weakness in some situations where we really want to examine unlabelled graphs. Classically this problem is controlled by a theorem of Wright ([B], Theorem IX.3) which says roughly that, unless M is extremely large or extremely small, the numbers of unlabelled and labelled graphs on n vertices with M edges, U_M and L_M respectively, satisfy

$$U_M \sim \frac{L_M}{n!} \text{ as } n \to \infty.$$

($U_M \geq L_M/n!$ clearly; the main work the other way is showing a.e. such graph has trivial automorphism group). A result for G_α follows from the theory linking the properties of G_M for values of M close to $n(n-1)\alpha/2$ to properties of G_α; see [B], Theorem II.2 for details. The link between G_M and our models is more tenuous, however, and we have no corresponding result.

Given n and k, the number of colours, we specify a k-vector of probabilities $\mathbf{s} = (s_1, s_2 ... s_k)^T$, where s_i is the probability that a vertex receives colour i, so that $s_i \geq 0 \, \forall i$ and $\sum_{i=1}^{k} s_i = 1$. We let Δ_k be the k-dimensional simplex of such vectors; the **support** of $\mathbf{s} \in \Delta_k$ is $\{i : s_i \neq 0\}$.

We also specify a k by k matrix P where p_{ij} is the probability that an edge between a vertex of colour i and one of colour j arises. Of course we must have the $p_{ij} \in [0, 1]$, and since our graphs are undirected P must be symmetric; these are the only restrictions on $\mathbf{s}$ and P.

Definition 1.1 *An* **RRC model of random graphs** *with parameters* $\mathbf{s}$ *and* P *is any model of random graphs obtained in the above way. We will denote it by* $\Gamma(n, k, \mathbf{s}, P)$.

We shall also often consider the **family** of such models as n varies, with k, $\mathbf{s}$ and P fixed. (It may well be interesting to study the situation where k, and hence $\mathbf{s}$ and P, vary with n in some structured way, but we shall not do so in this thesis. We will however sometimes deal with the case when, though $\mathbf{s}$ (and so also k) remain fixed, P varies with n).

As noted before, we will often compare the behaviour of some graph invariant in the $\Gamma(n, k, \mathbf{s}, P)$ model, or a family of such models, with a classical model (or family of models) where edges arise **independently** with the same average probability $\sum_{i=1 j=1}^{k} s_i s_j p_{ij} = \alpha$, so that any differences between the two models or families of models are due solely to the correlation structure in $\Gamma(n, k, \mathbf{s}, P)$. Since the events of interest are often that some collection of edges arises, we are often comparing the probability of what is in some sense the same event A in two different regimes; we shall usually describe it as $P\{A \text{ in } \Gamma(n, k, \mathbf{s}, P)\}$ or $P\{A \text{ in } G_\alpha\}$ to make it clear in what model the calculation is being carried out.

Let $N_i(\mathrm{S})$, for $S \subset V(G)$ be the number of vertices in S of colour i; when S is the whole of $V(G)$, we shall omit mention of it. We let $\mathbf{n} = (N_1, N_2 ... N_k)$. Of course $N_i \sim \mathrm{Bin}(n, s_i)$. Jointly, the N_i have a multinomial distribution;

$$P\{N_1 = n_1, N_2 = n_2, ..., N_k = n_k\} = \binom{n}{n_1, n_2, ... n_k} \prod_{i=1}^{k} s_i^{n_i}.$$

An obvious special case, where better results are often possible, is when $k = 2$, so that there are two colours, red and blue, and

$$P\{\text{v is red}\} = s, \; P\{\text{v is blue}\} = 1 - s, \; p_{11} = p, \; p_{22} = r, \; p_{12} = p_{21} = q.$$

Since such models often occur in this thesis, we introduce the special notation ${}_sG_{p,q,r}$ for them. Note that the first subscript after the G always refers to the red-red edge probability, the second to the red-blue probability, and the third to the blue-blue probability. We often specialise further to the case $s = 1/2$; then we omit s and write $G_{p,q,r}$. If the red-red and blue-blue probabilities are equal, we will write ${}_sG_{p,q}$ or $G_{p,q}$ rather than ${}_sG_{p,q,p}$ or $G_{p,q,p}$; thus if there are only two subscripts on the right, we have $p = r$. Since $G_{p,q}$ is the simplest RRC model other than the classical model, many of our results will be about it. We have

$$\alpha = s^2p + 2s(1-s)q + (1-s)^2r \text{ in } {}_sG_{p,q,r},$$

$$\alpha = \left(s^2 + (1-s)^2\right)p + 2s\,(1-s)\,q \text{ in } {}_sG_{p,q},$$

$$\alpha = \frac{p+2q+r}{4} \text{ in } G_{p,q,r} \text{ and } \alpha = \frac{p+q}{2} \text{ in } G_{p,q}.$$

We often write $P_{p,q}\{A\}$ or $P_\alpha\{A\}$ instead of $P\{A \text{ in } G_{p,q}\}$, or $P\{A \text{ in } G_\alpha\}$.

2 Manifestation of correlation structure in trees and cycles

2.1 TID models

We introduced $\Gamma(n, k, \mathbf{s}, P)$ in order to generate models with a non-trivial correlation structure so our first task is to investigate how these correlations are manifested in behaviour different from that of G_α. We shall see that, at least in families of RRC models, as n varies, correlation structure **is** always manifested, except in trivial cases; however, it will be convenient to delay the precise statement and proof until Theorem 2.17.

We saw in Chapter 1 that, by the definition of the corresponding classical model, the overall probability that a given edge arises is the same in both models. Thus the first question to address is when the probability of several edges all arising is the same as classically. We first note the intuitively obvious fact that collections of edges on disjoint vertex sets are independent. Note that the events we are dealing with here are just the events that some given edges arise; we say nothing about the other edges.

Lemma 2.1 *Suppose that, in some RRC model, A is the event that some collection of edges on vertex set U arises, B is the event that some collection of edges on vertex set V arises, and that U and V are disjoint so there are no common vertices or edges. Then A and B are independent events.*

Proof. Writing $c(U)$ and $c(V)$ for possible colourings of U and V

$$P\{A \cap B\} = \sum_{c(U),c(V)} P\{A \cap B \mid c(U), c(V)\} P\{c(U), c(V)\}.$$

Conditional on $c(U)$ and $c(V)$ the edges arise independently, so this is

$$\sum_{c(U),c(V)} P\{A \mid c(U), c(V)\} P\{B \mid c(U), c(V)\} P\{c(U), c(V)\}$$

$$= \sum_{c(U),c(V)} P\{A \mid c(U)\} P\{B \mid c(V)\} P\{c(U), c(V)\}$$

since whether or not the edges in A (respectively B) arise depends only on $c(U)$ (respectively $c(V)$). Since U and V are disjoint, they are coloured independently, so this is

$$\sum_{c(U),c(V)} P\{A \mid c(U)\} P\{B \mid c(V)\} P\{c(U)\} P\{c(V)\}$$

and carrying out the two summations this is $P\{A\}P\{B\}$ as required. •

Thus attention focuses on connected sets of edges. Often in graph theory the simplest connected graphs are **trees**, that is connected subgraphs with no cycles; a related notion is **forests**, graphs whose components are trees. (A particular type of tree we shall often consider is a **path**, a tree whose vertices can be renumbered $1, 2...n$ so that the edges are $i-(i+1)$ for each $1 \leq i \leq (n-1)$; the **length** of this path is $n-1$). Thus we ask when the probability that the edges of a tree arise in one of our models is the same as classically. Again note that we only ask whether or not the edges arise; we do not ask anything about the other edges. We make the following definition;

Definition 2.1 *An RRC model $\Gamma(n, k, \mathbf{s}, P)$ is TID (for* **tree indiscernible***) if and only if, for all trees T on $\leq n$ vertices, the probability that the edges of T arise is equal to the probability that they arise in the corresponding classical model.*

Theorem 2.2 *The model $\Gamma(n, k, \mathbf{s}, P)$ is TID if and only if, for all i such that $s_i \neq 0$, $\sum_{j=1}^{k} p_{ij}s_j = \alpha$.*

Proof. Any tree T can be built up sequentially. The first edge arises with probability α in both models. Thereafter, at each stage, we are at some vertex v and know its colour; by the definition of tree, the next vertex w has not yet occurred so its colour is unknown. Thus, taking T to be the path $1-2-3$, we require that $P\{v-w \mid c(v)\} = \alpha$ for all colourings $c(v)$ of v which have a non-zero probability of occuring and all vertices $v, w \in V(G)$, if the probability of this tree is to be the same as classically; that is, we need

$$\sum_{j=1}^{k} p_{ij}s_j = \alpha \ \forall 1 \leq i \leq k \text{ such that } s_i \neq 0.$$

Conversely if this condition holds, the model is TID, as required. •

Note that Theorem 2.2 shows that whether or not a family of models is TID can be tested on one particular tree, namely the path of length 2. Of course it does **not** imply that the distribution of the number of trees which arise is the same in the two models, as the joint existence of trees may be correlated in the new model even when they are not in the classical model. We discuss how the probability of trees compares with its classical value in non-TID models in section 2.4; for now, we draw corollaries of Theorem 2.2.

Corollary 2.3 *The probability of all the edges of a given forest F arising in G is the same in any TID model $\Gamma(n,k,\mathbf{s},P)$ and in the G_α.*

Proof. Since the various components in a forest have neither vertices nor edges in common, whether they arise or not is independent in both models, by Lemma 2.1. Thus, in both models, the probability that all the edges arise is just the product of the probabilities of the various components; the components of a forest are trees, and Theorem 2.2 shows that the probability of a tree arising is the same in the two models. The result follows. •

Theorem 2.2 also shows that there are lots of TID models. For example, if all the s_i are equal (so in particular none of them is zero) the model is TID if and only if all row (or column) sums of P are equal (to $k\alpha$); thus the matrix is a scaling of a stochastic symmetric matrix, and even the number a_n of symmetric permutation matrices (recall a permutation matrix has exactly one 1 in each row and column, all other entries being zero, so they are a subset of the symmetric stochastic matrices) satisfies $a_n \sim e^{-1/4}n^{n/2}e^{\sqrt{n}-n/2}/2$ by [VW, p128]. More precise information about the structure of the convex set of symmetric stochastic matrices may be harder to get since (for example) it is not true that every symmetric stochastic matrix is a convex combination of symmetric permutation matrices (in three dimensions, the four symmetric permutation matrices are

$$A=\begin{vmatrix}0&1&0\\1&0&0\\0&0&1\end{vmatrix}\quad B=\begin{vmatrix}0&0&1\\0&1&0\\1&0&0\end{vmatrix}\quad C=\begin{vmatrix}1&0&0\\0&1&0\\0&0&1\end{vmatrix}\quad D=\begin{vmatrix}1&0&0\\0&0&1\\0&1&0\end{vmatrix}$$

and then we have that $M=\lambda A+\mu B+\nu C+(1-\lambda-\mu-\nu)D$ is

$$\begin{vmatrix}1-\lambda-\mu & \lambda & \mu\\ \lambda & \mu+\nu & 1-\lambda-\mu-\nu\\ \mu & 1-\lambda-\mu-\nu & \lambda+\nu\end{vmatrix}$$

and thus we have the constraint that $m_{22}\geq m_{13}$; however the matrix

$$\begin{vmatrix}0.3&0.2&0.5\\0.2&0.4&0.4\\0.5&0.4&0.1\end{vmatrix}$$

though symmetric stochastic, does not satisfy this constraint.

Theorem 2.2 also allows us to completely classify when ${}_sG_{p,q,r}$ is TID; analogous results for models with several colours would be more complex.

Corollary 2.4 *${}_sG_{p,q,r}$ is TID if and only if $s = 0$, 1 or $(r-q)/(p+r-2q)$, or $p = r = q$. In particular, $G_{p,q}$ is always TID.*

Proof. The cases when $s = 0$ or 1 are easy, so we suppose $s \notin \{0,1\}$ Then using Theorem 2.2 the model is TID if and only if

$$s^2p + 2s(1-s)q + (1-s)^2 r = sp + (1-s)q = (1-s)r + sq \,(*)$$

$$\Leftrightarrow s(sp + (1-s)q) + (1-s)(sq + (1-s)r) = sp + (1-s)q = sq + (1-s)r$$

$$\Leftrightarrow s(p-q) = (1-s)(r-q).$$

Hence, if $p+r-2q \neq 0$ we get $s = (r-q)/(p+r-2q)$ and thus calculate the common value in $(*)$ to be $(rp-q^2)/(p-2q+r)$. If $p+r-2q = 0$ equation $(*)$ becomes $2s(q-r) + r = sp + (1-s)q = (1-s)r + sq$ and the previous equality implies (as $p+r-2q = 0$) $r = q$ and thus $p = q$ also.

Finally, to see that $G_{p,q}$ is TID, note that if $p = r$ and $p+r-2q \neq 0$ $(r-q)/(p+r-2q) = 1/2$ as required. If $p+r = 2q$, $p = r \Rightarrow q = p$ also. •

To close this section, we consider when, given a symmetric k by k matrix P with $p_{ij} \in [0,1]$, there exists $\mathbf{s} \in \Delta_n$ such that $\Gamma(n,k,\mathbf{s},P)$ is TID. We first note that we can always obtain degenerate TID models from old ones by adding further colours which all arise with probability 0, and expanding the matrix P by adding rows and columns, with any entries in $[0,1]$ we like, subject to the symmetry condition on P, to correspond to these colours. Thus we define $\Gamma(n,k,\mathbf{s},P)$ to be **non-degenerate** if $s_i \neq 0$ for all $1 \leq i \leq k$, and note that we can restrict attention to non-degenerate models; now the condition for tree indiscernibility in Theorem 2.2 is purely a system of linear equations in the p_{ij} and s_j.

Certainly, even then there need not be such an $\mathbf{s}$; for example if $r = 0.3$, $q = 0.4$, $p = 0.6$, $(r-q)/(p+r-2q) = -1 \notin [0,1]$ so no $G_{p,q,r}$ is TID, by Theorem 2.4. More generally, if some row of P is a multiple $\neq 1$ of another, $P\mathbf{s}$ cannot be a constant multiple of $\mathbf{1} = (1,1,...1)^T$ unless the multiplier is zero, which is impossible if $\mathbf{s}$ and P are positive. For the matrix with all entries equal to α however, any choice of $\mathbf{s}$ makes the model TID. We summarise the main restrictions in the following theorem.

Theorem 2.5 *Let $\mathbf{s}$ be a vector in Δ_k making $\Gamma(n,k,\mathbf{s},P)$ into a family of non-degenerate TID models. If $\mathbf{t}$ is any vector in Δ_n making $\Gamma(n,k,\mathbf{t},P)$ into a family of non-degenerate TID models, then $P\mathbf{t} = P\mathbf{s}$. Hence if P is*

invertible, there is at most one such $\mathbf{s}$*; there is such an* $\mathbf{s}$ *if and only if all components of* $P^{-1}\mathbf{1}$ *are* ≥ 0.

If P *is singular there is no such* $\mathbf{s}$ *if the equation* $P\mathbf{s} = \lambda\mathbf{1}$ *has no solutions with* $\mathbf{s} \in \Delta_k$ *and* $\lambda \geq 0$*; there is one solution if that equation is soluble and* $\exists \mathbf{v}$ *with* $P\mathbf{v} = \mathbf{0}$ *and* $\sum_{i=1}^{k} v_i = 0$*; otherwise there are infinitely many such* $\mathbf{s}$.

Proof. By Theorem 2.2, a non-degenerate model is TID if and only if we have $\sum_{j=1}^{k} p_{ij}s_j = \alpha \, \forall i$. This is if and only if $P\mathbf{s} = \lambda\mathbf{1}$ for some λ; the left to right implication is clear, and in the other direction, if $P\mathbf{s} = \lambda\mathbf{1}$, pre-multiplying both sides by $\mathbf{s}^T$ and using the facts that $\mathbf{s}^T P\mathbf{s} = \alpha$ and that $\sum_{i=1}^{k} s_i = 1$, we have $\lambda = \sum_{i,j=1}^{k} s_i s_j p_{ij} = \alpha$. Thus, if $P\mathbf{s} = \lambda_S\mathbf{1}$ and $P\mathbf{t} = \lambda_T\mathbf{1}$, we have, as $\sum_{i=1}^{k} s_i = 1$, that $\lambda_T = \mathbf{s}^T P\mathbf{t} = \mathbf{t}^T P\mathbf{s}$ by P symmetric; this is λ_S by the same argument, so $P\mathbf{s} = P\mathbf{t}$, and so if P is invertible $\mathbf{s} = \mathbf{t}$ as stated.

Otherwise P is singular. Then if $\mathbf{s} \neq \mathbf{t}$ both make the model TID, as $P\mathbf{s} = P\mathbf{t}$ by the previous paragraph we have $\mathbf{s} - \mathbf{t}$ is in the kernel of P and the sum of its components is zero. Conversely, if the kernel contains such a vector $\mathbf{z}$, adding a suitably small non-zero multiple of it to $\mathbf{s}$, all of whose components are positive since the model is non-degenerate, gives a distinct vector $\mathbf{t}$ which still has positive components summing to 1 and for which $P\mathbf{t} = \mu\mathbf{1}$ for some μ; then any convex combination of $\mathbf{t}$ and $\mathbf{s}$ will also have these properties, so there are indeed infinitely many choices of $\mathbf{s}$ making $\Gamma(n, k, \mathbf{s}, P)$ TID. •

2.2 Correlation structure in cycles in $G_{p,q}$.

In this section we consider cycles. It is clear that the argument of Theorem 2.2 fails for cycles, and we will show that this is not just an accident of the method of proof; cycles do show up the correlation structure, except in certain trivial cases. In this section, we deal with the case of $G_{p,q}$ where an explicit result is possible, before considering the general RRC model in section 2.3, where the result is less explicit. The fact that cycles are important in manifesting correlation structure in our models will be reinforced in chapters 3 and 4.

We first obtain the formula for the probability of a cycle in $G_{p,q}$. In the proof, for the sake of notational clarity, we describe the r-cycle whose edges join 1 and 2, 2 and 3, ..., $(r-1)$ and r, and r and 1 as $1 \to 2 \to ... \to r \to 1$, but this should not be misinterpreted as meaning that the edges are oriented.

Theorem 2.6

$$P_{p,q}\{1 \to 2 \to ... \to r \to 1\} = \alpha^r + \left(\frac{p-q}{2}\right)^r.$$

Proof. $P\{1 \to 2 \to ... \to r \to 1\} = P\{1 \to 2 \to ... \to (r+1) \mid c(1) = c(r+1)\}$. Set $P_r = P\{1 \to 2 \to ... \to r \mid c(1) = c(r)\}$, so that the probability of the cycle is P_{r+1}, and $Q_r = P\{1 \to 2 \to ... \to r \mid c(1) \neq c(r)\}$. By conditioning, and the fact that $G_{p,q}$ is TID by Corollary 2.4, we have

$$\alpha^{r-1} = P\{1 \to 2 \to ... \to r\} = \frac{P_r + Q_r}{2} \Rightarrow Q_r = 2\{\alpha^{r-1} - \frac{P_r}{2}\}\ (*)$$

Also, since the events that $1 \to 2 \to ... \to r$ arises and that $r \to (r+1)$ arises are independent conditional on $c(1)$, $c(r)$ and $c(r+1)$,

$$P_{r+1} = \frac{P_r p + Q_r q}{2} \Rightarrow P_{r+1} = P_r \frac{p-q}{2} + q\alpha^{r-1}.$$

by the formula $(*)$ for Q_r and some manipulative algebra. But the solution of the recurrence $x_r = ax_{r-1} + bc^{r-2}$ with initial condition $x_2 = d$ is

$$x_r = a^{r-2}d + cb\frac{c^{r-2} - a^{r-2}}{c-a} \text{ provided } a \neq c.$$

So here, we see that if $a = (p-q)/2 \neq c = \alpha$, as $d = p$ and $b = q$, that

$$P_{r+1} = p\left(\frac{p-q}{2}\right)^{r-1} + q\alpha\frac{\alpha^{r-1} - (\frac{p-q}{2})^{r-1}}{(\frac{p+q}{2} - \frac{p-q}{2})} \Rightarrow P_{r+1} = (p-\alpha)\left(\frac{p-q}{2}\right)^{r-1} + \alpha^r$$

which by some more manipulation is equal to the expression in the statement of the theorem. Otherwise $a = c$, that is $\frac{p-q}{2} = \frac{p+q}{2}$, so $q = 0$, and then the only way the cycle can arise is for all the vertices to be the same colour, which happens with probability $2^{-(n-1)}$; conditional on this, the cycle arises with probability p^n so the probability of the cycle is $\frac{p^n}{2^{n-1}}$ which is again easily seen to be equivalent to the expression in the statement of the theorem. •

Corollary 2.7

$$P_{p,q}\{1-2-...r \mid c(1) = c(r)\} = \alpha^{r-1} + \left(\frac{p-q}{2}\right)^{r-1} \text{ and}$$

$$P\{1-2-...r \mid c(1) \neq c(r)\} = \alpha^{r-1} - \left(\frac{p-q}{2}\right)^{r-1}.$$

Proof. The formula for P_r was proved explicitly in the theorem; the one for Q_r follows from it and $(*)$. •

The formulae in Corollary 2.7 will be used quite often in Chapter 3. In Theorem 8.3 we will see a slightly different argument which can be modified to give another proof of Theorem 2.6. A proof similar to Theorem 2.6 works rather more generally; for example, in a model with k equiprobable colours, all diagonal entries of P equal to p, and all off-diagonal entries q, it shows

$$P\{\text{an } r\text{-cycle}\} = \left(\frac{p+(k-1)q}{k}\right)^r + (k-1)\left(\frac{p-q}{k}\right)^r$$

but as Theorem 2.12 will subsume this result, we omit the proof.

Corollary 2.8 *The expected number of cycles of even length is larger in $G_{p,q}$ than G_α for all r. The expected number of cycles of odd length is larger in $G_{p,q}$ than in G_α when $p > q$ and smaller when $p < q$. In particular, the probability is the same as classically if and only if $p = q$.* •

This can be thought of informally as reflecting the fact that if $q > p$ the graph is becoming more like a bipartite graph with classes the reds and the blues. (Recall a bipartite graph can only have cycles of even length).

Note that these calculations for trees and cycles depend heavily on the good exchangeability properties of our model. With predetermined numbers of reds and blues, the probability of a tree or cycle would depend on the colours of previous vertices making the formulae more complex.

2.3 The probability of a cycle in more general models

We now find the probability of a cycle in any $\Gamma(n, k, \mathbf{s}, P)$. We first define a non-negative symmetric matrix Q and a non-negative vector $\mathbf{v}$ by

$$q_{ij} = \sqrt{s_i}p_{ij}\sqrt{s_j} \text{ and } \mathbf{v}_i = \sqrt{s_i} \text{ for } 1 \le i, j \le k.$$

so that $\mathbf{v}^T Q\mathbf{v} = \alpha$ and $(\mathbf{v}, \mathbf{v}) = 1$, where $(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^k x_i y_i$ is the usual inner product for k-dimensional real vectors. We shall often make use of the elementary fact that Q has k real eigenvalues (counted with multiplicity), and that there is an orthonormal basis of eigenvectors of Q. The trick of working with Q rather than P is, since it depends on taking the square roots of the s_i, rather unnatural from a probabilistic point of view, but it clearly works.

The next theorem summarises the basic facts from the Perron-Frobenius theory of non-negative matrices, which we shall use at various points in what follows. Note that we deal mostly with symmetric matrices, making the distinction between left and right eigenvectors redundant.

Theorem 2.9 *Let A be a non-negative k by k matrix. Suppose that for every pair i, j of indices, $1 \leq i, j \leq k$ there is a positive integer $m(i,j)$ with $A^{m(i,j)}$ having its (i,j) entry positive; we say A is* **irreducible**. *Then there exists an eigenvalue λ (the maximal, or Perron-Frobenius eigenvalue) of A with the following properties;*

1. λ is real and positive.

2. With λ can be associated positive right and left eigenvectors of A.

3. $\lambda \geq \mid \mu \mid$ for any other eigenvalue $\mu \neq \lambda$. If the stronger condition than irreducibility, that A is **primitive** *holds (that is, there exists $k > 0$ with all entries of A^k positive) holds, then $\lambda > \mid \mu \mid$ for all other eigenvalues μ.*

4. The eigenvectors associated with λ, for any irreducible A, are unique (up to constant multiples).

5. For any irreducible A, λ is between the minimal and maximal row sums of A.

Proof. [Se] pp 1-6 and 20. •

We start by obtaining a general formula for the probability of an r-cycle.

Theorem 2.10

$$P\{1 \to 2 \to \ldots \to r \to 1\} = \sum_{i=1}^{k} \lambda_i^r \text{ in } \Gamma(n, k, \mathbf{s}, P)$$

where λ_i are the k eigenvalues of Q counted with multiplicity.

Proof. Conditioning on the colours of the vertices of the r-cycle

$$P\{1 \to 2 \to \ldots \to r \to 1\} = \sum_{i_1,\ldots i_r=1}^{k} s_{i_1} s_{i_2} \ldots s_{i_r} p_{i_1 i_2} p_{i_2 i_3} \ldots p_{i_r i_1}$$

$$= \sum_{i_1,\ldots i_n=1}^{k} q_{i_1 i_2} q_{i_2 i_3} \ldots q_{i_r i_1}$$

using the definition of Q. We now sum out the variables one by one, each such summation corresponding to a matrix multiplication, obtaining

$$\sum_{i_1=1}^{k} (Q^r)_{i_1 i_1} = \mathrm{tr}\,(Q^r) = \sum_{i=1}^{k} \lambda_i^r$$

where tr denotes the trace; the last equivalence is simple linear algebra. •

We first check that this result includes the generalisation of Corollary 2.7 mentioned after that result, using the following easy lemma.

Lemma 2.11 *Let I be the k by k identity matrix, and J the k by k matrix all of whose entries are equal to 1. Then the eigenvalues of $P = (p-q)\,I + qJ$ are $p + (k-1)\,q$ with multiplicity 1 and $p - q$ with multiplicity $(k-1)$.*

Proof. Clearly $\mathbf{1}$ is an eigenvector with eigenvalue $p + (k-1)q$. As P is symmetric it has an orthogonal basis of eigenvectors, so any other eigenvector $\mathbf{v}$ is perpendicular to $\mathbf{1}$; thus $J\mathbf{v} = 0$ and so $P\mathbf{v} = (p-q)\mathbf{v}$ as required. •

Theorem 2.12 *Suppose we have a model with k equiprobable colours, with all diagonal entries of P being p, and all off-diagonal entries q. Then*

$$P\{\textit{an } r\textit{-cycle}\} = \left(\frac{p + (k-1)q}{k}\right)^r + (k-1)\left(\frac{p-q}{k}\right)^r.$$

Proof. Since the s_i are all equal, we will have $Q = \frac{1}{k}P$. Now $P = (p-q)I + qJ$ and so the result follows from Lemma 2.11 and Theorem 2.10. •

It is interesting to note that the error term here is (for large r) smaller when $k > 2$ than when $k = 2$; informally speaking, having many colours to choose from dilutes the correlation structure.

The natural analogue for cycles of the notion of tree indiscernibility is

Definition 2.2 *An RRC model $\Gamma(n, k, \mathbf{s}, P)$ is* **cycle indistinguishable** *(CID) if and only if the probability of every cycle (of whatever length) in that model is equal to the probability of that cycle in the corresponding classical model.*

Again, note that we are talking about the probability that the cycle arises, possibly with some other edges, not that the cycle arises and nothing else.

We now show that the definition is a damp squib; any CID model with $n \geq 8$ is essentially the same as the corresponding classical model. The proof proceeds by first understanding the spectrum of Q in a CID model.

Theorem 2.13 *Suppose we have an RRC model $\Gamma(n,k,\mathbf{s},P)$ with $n \geq 8$. Then $\Gamma(n,k,\mathbf{s},P)$ is CID if and only if Q has one eigenvalue equal to α and all other eigenvalues are 0.*

Proof. The case when $Q = 0$, or equivalently $\alpha = 0$, is trivial, so we assume $Q \neq 0$ and so $\alpha \neq 0$ for the rest of the proof. As the model is CID, $\alpha^r = \sum_{j=1}^{k} \lambda_j^r$ for all $r \leq n$ by Theorem 2.10. In particular, $\alpha^4 = \sum_{i=1}^{k} \lambda_i^4$ and $\alpha^8 = \sum_{i=1}^{k} \lambda_i^8$. Note that all terms in both of these relations are real and non-negative. Squaring the first relation, and equating the two expressions for α^8, we have that the sum of all the cross-terms $\lambda_i^4 \lambda_j^4$ $(i \neq j)$ must be zero; as the cross-terms are themselves non-negative, each cross-term must be zero, and this implies that all but at most one of the λ_i^4 must be zero; thus all but at most one of the λ_i must be zero; then the remaining one must be α.

The converse is immediate from Theorem 2.10. •

In fact, the additional restrictions on Q enable us to say quite a bit more.

Lemma 2.14 *If $\Gamma(n,k,\mathbf{s},P)$ is a CID model with $n \geq 8$, the eigenvector of Q with eigenvalue α (whose existence and uniqueness follow from Theorem 2.13) is $\mathbf{v} = (\sqrt{s_1}, \sqrt{s_2}, \ldots \sqrt{s_k})^T$.*

Proof. Again the case when $Q = 0 \Leftrightarrow \alpha = 0$ is trivial, so we assume $\alpha \neq 0$ for the rest of the proof. As Q is symmetric, it has a basis of eigenvectors $\{\mathbf{e}_i\}$ for $1 \leq i \leq k$ which are orthonormal with respect to the standard inner product $(,)$; let $\mathbf{e}_1$ be the eigenvector with eigenvalue α. If $\mathbf{v} = \sum_{j=1}^{k} \mu_j \mathbf{e}_j$, then $Q\mathbf{v} = \mu_1 \alpha \mathbf{e}_1 \Rightarrow \alpha = (Q\mathbf{v}, \mathbf{v}) = \mu_1^2 \alpha \Rightarrow \mu_1^2 = 1$ since $\alpha \neq 0$. Also $(\mathbf{v}, \mathbf{v}) = 1 \Rightarrow \sum_{j=1}^{k} \mu_j^2 = 1 \Rightarrow \mu_j = 0$ for $j \geq 2$ so $\mathbf{v} = \mu_1 \mathbf{e}_1$ is an eigenvector of Q with eigenvalue α, as required. •

Lemma 2.15 *Let $\Gamma(n,k,\mathbf{s},P)$ be a CID model with $n \geq 8$. If $Q \neq 0$, Q consists (after re-numbering the colours) of an r by r block of positive entries $q_{ij} = \lambda_i \lambda_j q_{11}$ for $1 \leq i,j \leq r \leq k$, with zeroes in the other $k^2 - r^2$ positions; also, all the entries of the r by r block are positive.*

Proof. If $\alpha = 0$, then $Q = 0$ and the result is trivial. Otherwise $\alpha > 0$. By Theorem 2.13, Q has rank 1 so any two rows are linearly dependent. Thus if two rows both contain a non-zero entry, either is a non-zero multiple of the other. Renumber the colours so that the rows which do not consist entirely of zeroes are the first $r \leq k$ rows. Thus (by symmetry) any non-zero entries are

in the top r by r submatrix of Q, which we shall call R; R inherits symmetry from Q, each row of R has at least one non-zero entry, and the i-th row of R is a non-zero multiple λ_i of the first row; thus by symmetry also the i-th column is λ_i times the first column. Thus $q_{ij} = \lambda_i q_{1j} = \lambda_i \lambda_j q_{11}$ and so, since some q_{ij}, $1 \leq i, j \leq r$ is non-zero, q_{11} is non-zero and thus all the $q_{ij} > 0$, as claimed. •

We can now pull strings together.

Theorem 2.16 *Let $\Gamma(n, k, \mathbf{s}, P)$ be a CID model with $n \geq 8$. Then there exists $r \leq k$ such that after renumbering the colours, the support of $\mathbf{s}$ is the first r components and P has an r by r block with all elements equal to α in the top left-hand corner, the other $k^2 - r^2$ entries being arbitrary (subject to P being symmetric with entries in $[0, 1]$). In particular, the probability of any set of m edges arising is α^m irrespective of the colouring, so that the model is essentially a classical one.*

Proof. Again if $\alpha = 0$, $Q = 0$ and everything is trivial; thus we assume $\alpha \neq 0$ and so $Q \neq 0$. By Lemma 2.15, after renumbering the colours, Q consists of an r by r block, $r \leq k$, where there exist $\lambda_i > 0$ such that $q_{ij} = \lambda_i \lambda_j q_{11} > 0$ for $1 \leq i, j \leq r$ and has its other $k^2 - r^2$ entries zero. Note that the trace of this block must be α since this is the sum of the eigenvalues of the block, so

$$\sum_{j=1}^{r} \lambda_j^2 q_{11} = \alpha \, (*).$$

Also, $\mathbf{v}$ (renumbered at the same time as Q obviously) is the eigenvector of Q with eigenvalue α by Lemma 2.14. Using the fact that $\mathbf{v}^T Q \mathbf{v} = \alpha$, we have

$$\sum_{i,j=1}^{r} v_i \lambda_i q_{11} \lambda_j v_j = \alpha \Rightarrow \sum_{i,j=1}^{r} \lambda_i v_i \lambda_j v_j = \frac{\alpha}{q_{11}}.$$

By the Cauchy-Schwarz inequality, we thus have

$$\frac{\alpha}{q_{11}} = (\sum_{j=1}^{r} v_j \lambda_j)^2 \leq \sum_{j=1}^{r} v_j^2 \sum_{j=1}^{r} \lambda_j^2 \leq \sum_{j=1}^{k} v_j^2 \sum_{j=1}^{r} \lambda_j^2 = \frac{\alpha}{q_{11}}.$$

Here we use the fact that $\sum_{j=1}^{k} v_j^2 = 1$ by the definition of $\mathbf{v}$ together with the fact that $\mathbf{v}_j = 0$ for $j > r$, and the fact $(*)$ above to evaluate the second sum. Thus we have equality in Cauchy-Schwarz, which implies that $\lambda_j = c v_j$

for all $1 \leq j \leq r$, where c is some constant; in fact, $c^2 q_{11} = \alpha$ by use of $(*)$ again. Thus

$$v_i p_{ij} v_j = q_{ij} = \lambda_i \lambda_j q_{11} = v_i c^2 q_{11} v_j = v_i \alpha v_j \, \forall 1 \leq i, j \leq r$$

and as $v_i \neq 0$ for $1 \leq i, j \leq r$ we see $p_{ij} = \alpha$ for these values of i and j (the others are essentially arbitrary) and now all claims have been proven. •

That being CID is a highly demanding property is in keeping with the remarks on the central role of cycles in showing up correlation structure made after Corollary 2.7. Note that in particular, this shows that CID implies TID; we will reprove this in section 2.4. Of course there are TID models which are not CID; an example is $G_{p,q}$ with $p \neq q$, by Theorem 2.6.

Theorem 2.16 makes it easy for us to show that correlation structure is always manifest in families of RRC graphs except in the circumstances described in the statement of Theorem 2.17, which we will henceforth refer to as **trivial** cases (note that degenerate and trivial are distinct notions!).

Theorem 2.17 *Let $\Gamma(n, k, \mathbf{s}, P)$ be an RRC model with $n \geq 8$. Then the edges arise independently with constant probability α if and only if all entries of P which correspond to colours arising with non-zero probability are equal to α.*

Proof. If the edges arise independently with probability α, then the probability of any cycle is equal to its classical probability, so the model is CID and hence of the form stated by Theorem 2.16. On the other hand, if the model is as stated, we have, for any collection of edges, that the probability that they all arise is equal to the probability that they arise conditional on all vertices taking one of the colours which arise with non-zero probability, by Bayes' theorem, since the probability of any vertex taking another colour is zero and the vertices are coloured independently. Since after this conditioning the edges do arise independently with probability α, the edges do indeed arise independently with probability α. •

It may be possible to get pathological behaviour for $n \leq 7$; this is certainly possible for $n = 3$.

One aspect of Theorem 2.10 is that it gives the asymptotic rate of decay for the probability of the cycle;

Theorem 2.18 *Suppose C_r denotes an r-cycle, and $\lambda_1 \geq \ldots \geq \lambda_k$ are the eigenvalues of Q (which are all real as Q is symmetric). Then*

$$\lim_{r \to \infty} \frac{\log P\{C_{2r}\}}{2r} = \log(\lambda_1) \text{ and } \lim_{r \to \infty} \frac{\log P\{C_{2r+1}\}}{2r+1} = \log(\mu)$$

where μ is the largest eigenvalue of Q for which $-\mu$ is not an eigenvalue with the same multiplicity (or zero if the spectrum of Q is symmetric about the origin).

Proof. By Theorem 2.10

$$\lim_{r\to\infty} \frac{\log P\{C_r\}}{r} = \lim_{r\to\infty} \frac{\log(\sum_{i=1}^{k} \lambda_i^r)}{r} \quad (*)$$

$$= \lim_{r\to\infty} \left(\frac{\log(\lambda^r)}{r} + \frac{\log(1 + \sum_{i=2}^{k} (\lambda_i/\lambda)^r)}{r}\right).$$

If r is even, this is

$$\log(\lambda) + \lim_{r\to\infty} \frac{\log(1+B)}{r}$$

where $0 \leq B \leq k-1$, and the limit term goes to zero as $r \to \infty$ giving the claim.

Otherwise r is odd. Then for each eigenvalue η for which $-\eta$ is an eigenvalue of the same multiplicity, the rth powers of η and $-\eta$ cancel in the expression $(*)$. For the largest eigenvalue μ for which $-\mu$ is not an eigenvalue with equal multiplicity, μ must have greater multiplicity than $-\mu$ (consider a large odd value of r in Theorem 2.10 and use the fact that the probability of an r-cycle is non-negative). If the multiplicity of μ is n_1 and that of $-\mu$ is $n_2 < n_1$, we have by $(*)$

$$\lim_{r\to\infty} \frac{\log(\sum_{i=1}^{k} \lambda_i^r)}{r} = \lim_{r\to\infty} \left(\frac{\log(n_1 - n_2)}{r} + \frac{\log(\mu^r)}{r} + \frac{\log(1 + \sum(\lambda_i/\mu)^r/(n_1 - n_2))}{r}\right)$$

$$= \log(\mu) + \lim_{r\to\infty} \left(\frac{\log(n_2 - n_1)}{r} + \frac{\log(1 + \sum(\lambda_i/\mu)^r/(n_1 - n_2))}{r}\right)$$

where the sum is now over only those λ_i whose modulus is less than μ; hence by the same argument as before, the error term goes to zero, and the claim is now proven. •

The case where the spectrum of Q is symmetric about the origin can happen, for example in $G_{0,q}$ where as observed before the probability of a cycle of odd length is indeed zero. Of course if Q is primitive, $\lambda_1 = \mu$.

2.4 Are trees more likely to arise than classically?

In the previous sections we discussed under what circumstances the probability that a tree or a cycle arises is the same as classically. We now ask how such probabilities compare with the classical value when they are not equal to it, and in particular when they are greater than or less than classically. In this section we consider this problem for trees, and in the next section we give results on the question for cycles, which will depend on one of the results we prove in this section.

Initial calculations of the probabilities of some trees on small numbers of vertices in some simple RRC models suggested that they are always at least as likely to arise as classically, and this, after proving the result in some special cases, eventually led us to make the following conjecture.

Conjecture. Let $\Gamma(n, k, \mathbf{s}, P)$ be any RRC model, and let T be a tree. Then the probability that all edges of T arise is at least as large as in the corresponding classical model.

We do not offer the conjecture with any great confidence; we have no very firm heuristic to make us believe it, merely the various special cases we have proven. However, even if the conjecture is in fact false, the question of when trees are more or less likely to arise than classically is likely to be of interest.

We again emphasise that we are talking about the probability that all the edges of the tree are present, not the probability that these edges arise and that no others do; the latter probability seems to be harder to talk about in general. For example, even in the simple case of a graph on three vertices, let A be the event that the edge $1-2$ arises and neither $1-3$ nor $2-3$ does, and B be the event that the edges $1-2$ and $1-3$ arise but that $2-3$ does not arise; then a short calculation conditioning on the colours of the vertices shows that

$$P\{A \text{ in } G_{p,q}\} - P\{A \text{ in } G_\alpha\} = (\frac{p-q}{2})^3 \quad \text{but}$$

$$P\{B \text{ in } G_{p,q}\} - P\{B \text{ in } G_\alpha\} = -(\frac{p-q}{2})^3$$

which is positive for $p > q$ in one case and negative for $p > q$ in the other.

We first show that an obvious naive approach to proving the conjecture does not work, at least in simplistic form. It is tempting to believe that we

can argue along the following lines; consider some vertex v of degree 1 (recall that the degree of a vertex is the number of vertices adjacent to it) in the tree T (any tree has at least two vertices of degree 1), and let T' denote T with v removed; then, considering the colour of the vertex w adjacent to v in T, we have that

$$P\{T\} = \sum_{i=1}^{k} \left(\sum_{j=1}^{k} p_{ij} s_j \right) s_i P\{T' \mid c(w) = i\}$$

and it seems heuristically reasonable (since getting off to a good start ought, other things being equal, to leave one ahead at the end) that the sequence $P\{T' \mid c(w) = i\}$ should be increasing with the $\sum_{j=1}^{k} p_{ij} s_j$; if this were true, we would be done via the following standard inequality.

Lemma 2.19 *Suppose* $s_i \geq 0$ *with* $\sum_{i=1}^{k} s_i = 1$ *and* $0 \leq a_1 \leq a_2 \leq ...a_k$ *and* $0 \leq b_1 \leq b_2 \leq ...b_k$ *are increasing sequences. Then*

$$\sum_{i=1}^{k} s_i a_i b_i \geq \sum_{i=1}^{k} s_i a_i \sum_{i=1}^{k} s_i b_i.$$

If the a_i *are increasing and the* b_i *are decreasing, we get*

$$\sum_{i=1}^{k} s_i a_i b_i \leq \sum_{i=1}^{k} s_i a_i \sum_{i=1}^{k} s_i b_i.$$

Proof. This can be proven directly; in addition, it is immediate from the FKG inequality (Theorem 3.14). •

However the heuristic step in the above argument is false. We can see this by considering the case of a **star**, that is a graph on n vertices in which $n-1$ vertices have degree 1 and the other has degree $n-1$; clearly stars are trees. Then, letting v be the vertex of degree $n-1$ and w be one of the $n-1$ vertices of degree 1, in ${}_sG_{p,q,r}$ with $q = 1$, $p > r$ but both much smaller than q, we have, conditioning on $c(v)$, that

$$P\{T \mid c(w) = \text{red}\} - P\{T \mid c(w) = \text{blue}\}$$

$$= sp\,(1 - s + sp)^{n-2} + (1 - s)\,(s + (1 - s)\,r)^{n-2}$$

$$-s\,(1 - s + sp)^{n-2} - (1 - s)\,r\,(s + (1 - s)\,r)^{n-2}$$

$$= s(p-1)(1-s+sp)^{n-2} + (1-s)(1-r)(s+(1-s)r)^{n-2}$$
$$= \frac{1}{2}\left(\frac{p+1}{2}\right)^{n-2}\left(p-1+(1-r)\left(\frac{1+r}{1+p}\right)^{n-2}\right) \text{ if } s=\frac{1}{2}$$

and because $p > r$ this will be less than zero for sufficiently large n. Thus, using the second part of Lemma 2.19, we see that in fact the probability of the tree consisting of the above, with a further edge $w-u$, is less than the product of the probability of the star and the probability of the edge. This is of course not itself inconsistent with the conjecture, and indeed Theorem 2.24 will prove the conjecture for stars, but it does suggest that if the conjecture is true, it will not be for entirely straightforward reasons; in the language of dynamic programming, the problem is not a one-step problem.

(One might note that whilst the condition that the a_i and b_i are both ordered the same way is sufficient for the inequality in the first half of Lemma 2.19, it is clearly not necessary, and ask what work has been done on extensions of Lemma 2.19; but the only result we are aware of in this direction [SS] is not helpful, partly because it is only stated for the case of equiprobable colours, but mainly because the condition given for the inequality to hold cannot be checked without much more detailed knowledge of the probabilities of trees conditional on the colour of one vertex than we have at present).

We first prove the conjecture in the important special case of a path. To do so, we first obtain a formula for the probability of a path in a general RRC model, using ideas similar to those in Theorem 2.10.

Theorem 2.20 *In an RRC model* $\Gamma(n,k,\mathbf{s},P)$, *with* Q *and* $\mathbf{v}$ *as before,*

$$P\{\textit{a path of length } (r-1)\} = \sum_{i_1,i_r=1}^{k} \left(Q^{r-1}\right)_{i_1 i_r} \sqrt{s_{i_1} s_{i_r}} = \mathbf{v}^T Q^{r-1}\mathbf{v}.$$

Proof. The probability is

$$\sum_{i_1,i_2\ldots i_r=1}^{k} s_{i_1} s_{i_2}\cdots s_{i_r} p_{i_1 i_2} p_{i_2 i_3}\cdots p_{i_{r-1} i_r}$$

which, introducing Q as in the proof of Theorem 2.10 for cycles is

$$\sum_{i_1,i_2\ldots i_r=1}^{k} \sqrt{s_{i_1}}\, q_{i_1 i_2} q_{i_2 i_3}\cdots q_{i_{r-1} i_r} \sqrt{s_{i_r}}$$

and performing matrix multiplication again as in the proof of Theorem 2.10 this is indeed

$$= \sum_{i_1,i_r=1}^{k} \sqrt{s_{i_1}} \left(Q^{r-1}\right)_{i_1 i_r} \sqrt{s_{i_r}} = \mathbf{v}^T Q^{n-1} \mathbf{v}. \bullet$$

In the case when the colours are equiprobable, the formula reduces to

$$\sum_{i,j=1}^{k} \left(\frac{1}{k}\right)^r \left(P^{r-1}\right)_{i,j}$$

which again is easily checked to agree with the known formula in $G_{p,q}$. However this formula is less tidy than Theorem 2.10 since the lack of a final edge back to the start leaves the weight $\sqrt{s_{i_1} s_{i_r}}$ in the sum. More importantly, this argument applies only to paths, not all trees.

The other main tool we shall use is the following inequality, motivated by a problem in genetics, and due to Mulholland and Smith.

Theorem 2.21 *Let* $\mathbf{w}$ *be a non-negative* k*-vector and* A *a non-negative* k *by* k *symmetric matrix. Then*

$$\mathbf{w}^T A^n \mathbf{w} \left(\mathbf{w}^T \mathbf{w}\right)^{n-1} \geq \left(\mathbf{w}^T A \mathbf{w}\right)^n$$

with equality if and only if $n = 1$ *or* $\mathbf{w}$ *is an eigenvector of* A.

Proof. [MS]. •

Theorem 2.22 *The probability that the path* $1 - 2 - \ldots - n$ *arises is always at least as large in* $\Gamma(n, k, \mathbf{s}, P)$ *as classically. For* $n > 2$*, if there is equality,* $\mathbf{v}$ *must be an eigenvector of* Q.

Proof. The probability of the path is, by Theorem 2.20

$$\mathbf{v}^T Q^{n-1} \mathbf{v}$$

where $\mathbf{v} = (\sqrt{s_1}, \ldots \sqrt{s_k})^T$ as usual. Taking the matrix A in the Mulholland-Smith theorem to be Q and the vector $\mathbf{w}$ to be $\mathbf{v}$, we rapidly see that all the conditions of the theorem are satisfied, and so the above is

$$\geq \frac{\left(\mathbf{v}^T Q \mathbf{v}\right)^{n-1}}{\left(\mathbf{v}^T \mathbf{v}\right)^{n-2}} = \frac{\alpha^{n-1}}{1} = \alpha^{n-1}$$

with equality if and only if $\mathbf{v}$ is an eigenvector of Q. •

In Chapter 3 we use the fact that the path $1-2-3$ is at least as likely as classically to show that the variance of the number of edges is no less than classically in RRC models; this special case can easily be proven directly using the Cauchy-Schwarz inequality. We discuss the implications of Theorem 2.22 for probabilities of cycles in section 2.5. It also gives new insight on when a model is TID;

Lemma 2.23 *A model is TID if and only if* $\mathbf{v}$ *is an eigenvector of* Q *(when its eigenvalue is automatically* α*).*

Proof. If the model is TID, the probability of a path of length r is α^r, so $\mathbf{v}$ is an eigenvector of Q by Theorem 2.20 and Theorem 2.21. Conversely, if $\mathbf{v}$ is an eigenvector of Q, $Q\mathbf{v} = \mu\mathbf{v}$ for some μ; multiplying on the left by $\mathbf{v}^T$ we deduce $\mu = \alpha$, and so writing the relationship out, $\sum_{j=1}^{k} \sqrt{s_i} p_{ij} s_j = \alpha\sqrt{s_i}$ so if $s_i \neq 0$, we have $\sum_{j=1}^{k} p_{ij} s_j = \alpha$, which is exactly the condition in Theorem 2.2 for tree-indiscernibility. •

In particular, this makes it clear that CID implies TID, by Lemmas 2.23 and 2.14 although of course Theorem 2.16 supersedes this result.

We next give the promised proof of the conjecture for stars.

Lemma 2.24 *Let* T *be a star. Then* $P\{T \text{ in } \Gamma(n,k,\mathbf{s},P)\} \geq P\{T \text{ in } G_\alpha\}$.

Proof. Conditioning on the colour of the vertex of high degree, we see $P\{T\} = \sum_{i=1}^{k} s_i a_i^{n-1}$ where $a_i = \sum_{j=1}^{k} p_{ij} s_j$ is the probability of an edge conditional on the colour of one end being i, and so is non-negative. By the convexity of the function $x \to x^r$ for $x \geq 0$ and $r \geq 1$, this is at least $(\sum_{i=1}^{k} s_i a_i)^{n-1} = \alpha^{n-1}$ which is the classical probability of the star. •

This argument works for any tree T consisting of several copies of some other tree T' all joined (in the same way) to a single central vertex (so that there is symmetry present), provided that the probability of T' is at least as large as classically (for example, if T' is a path).

Our last piece of evidence for the conjecture is a theorem of Kingman, which is a simplification and generalisation of an earlier proof by Atkinson, Watterson and Moran of a matrix inequality; Kingman himself later generalised his result to a version concerning Radon-Nikodym derivatives rather than partial averages. We change his notation slightly to avoid confusion with our use of the term α; in our application the numbers $p_i, q_i, ..$ will all be equal to s_i.

Theorem 2.25 *Let $a_{ijk...}$ be a set of non-negative numbers, and denote the set of indices by κ. If we have non-negative numbers p_i, $q_j...$ with $\sum p_i = \sum q_j = ... = 1$, and $\kappa = \beta \cup \beta'$ is a partition of κ into two subsets, let $p_i q_j ... = P_\kappa = Q_\beta Q'_{\beta'}$ and define the* **partial average** *of a_κ over β' to be*

$$A_\kappa(\beta) = \sum_{\beta'} a_\kappa Q'_{\beta'}$$

so that, if for example $\kappa = \{i,j\}$ and $\beta = \{i\}$, $A_\kappa(\beta) = \sum_{k=1} a_{ik} s_k$, and if β is empty, $A_\kappa(\beta) = \alpha$. Then, if the $A_\kappa(n)$, for $n = 1, 2, ...$ are some partial averages of the a_α, and λ_n are non-negative, we have

$$\sum_\kappa a_\kappa P_\kappa \prod_n A_\kappa^{\lambda_n}(n) \geq \alpha^{1+\sum \lambda_n}.$$

Proof. [Ki] •

This theorem allows us to show that the probability of a tree consisting of two stars, which we then join by adding an edge between the two centres, is at least as large as classically. More generally it shows that the probability of some tree, with copies of itself attached at various points, is at least as large as classically, taking $a_{ijk..lm} = p_{ij} ... p_{lm}$ to be the product of the probabilities of the various edges in the tree conditional on their colourings, and is thus somewhat more general than the result about stars. The above suggests (informally speaking) that any counterexample to the conjecture must have a fairly high degree of asymmetry.

These various proof techniques amongst them are easily seen to imply that any counterexample to the conjecture must have at least six vertices. The simplest case which is not covered by any of the results is when the tree has one vertex of degree 3, with one vertex of degree 1 joined to it, and two paths of length 2. However one can check by a computation that the probability of this tree arising in ${}_sG_{p,q,r}$ is at least as large as classically.

Another obvious testing ground for the conjecture is when the matrix P is of a special form. Suppose for example P has zeroes on the diagonal and ones elsewhere. Then the probability that the tree arises is clearly equal to the probability that the colouring is a proper colouring, in the sense of chromatic numbers. In the case when the colours are equiprobable, the model is TID, and so, since $\alpha = 1 - 1/k$ clearly, this allows us to recover the fact that the number of proper colourings (which is k^n times the probability that a colouring is proper) is equal to $k(k-1)^{n-1}$, though of course this fact is

more easily obtained in other ways. When the colours are not necessarily equiprobable, so that $\alpha = 1 - \sum_{j=1}^{k} s_j^2$, the situation seems harder to understand in general. However for $k = 2$ it yields easily to the arithmetic-mean geometric-mean inequality. Indeed if $k = 2$ we can write $\mathbf{s} = (s, 1-s)^T$ and consider whether some fixed vertex v is red or blue. Then note that the colours must alternate for the tree to arise, and so the probability of the tree is $s^a(1-s)^b + s^b(1-s)^a$, where a is the number of vertices at even (including zero) distance from v and b the number of vertices whose distance from v is odd, so that $a + b = n$. Thus

$$\begin{aligned} & P\{T \text{ in } \Gamma(n,k,\mathbf{s},P)\} - P_\alpha\{T\} \\ &= s^a(1-s)^b + s^b(1-s)^a - (1 - s^2 - (1-s)^2)^{a+b-1} \\ &\geq 2\sqrt{s^{a+b}(1-s)^{a+b}} - (2s(1-s))^{a+b-1} \text{ by AM-GM inequality} \\ &= 2s^{(a+b)/2}(1-s)^{(a+b)/2}(1 - 2^{a+b-2}s^{(a+b)/2-1}(1-s)^{(a+b)/2-1} \\ &= 2s^{(a+b)/2}(1-s)^{(a+b)/2}(1 - (4s(1-s))^{(a+b)/2-1}) \end{aligned}$$

and this is non-negative since $4s(1-s) \leq 1$ for $s \in [0,1]$ (either by calculus or further use of the AM-GM inequality), and $(a+b)/2$ is positive.

It is natural to ask if we can deal with the case when T is an m-ary tree, that is T has a vertex (the ancestor), joined to m other vertices (the first generation), and for each vertex of the first generation there are m other neighbours (the second generation), and so on for n generations. The cases of one and two generations are covered by the above arguments; in general, it is very easy to write down the following relations for the $f_{n,i}$, the probability that an m-ary tree with n generations arising conditional on the root having colour i;

$$f_{n,i} = \left(\sum_{j=1}^{k} p_{ij} s_j f_{n-1,j} \right)^m$$

simply by conditioning on the colours of the vertices adjacent to the root; this recurrence relation (whose initial conditions are $f_{0,i} = 1 \forall i$ and $f_{1,i} = \left(\sum_{j=1}^{k} p_{ij} s_j\right)^m$) seems however to be intractable, at least partly because it is non-linear.

Since we cannot at present prove the full conjecture, we make two speculative remarks which may provide some insight into an eventual solution. Firstly, note that the proofs of Theorem 2.24 and Theorem 2.25 both use a convexity inequality, and Theorem 2.22 can be shown to be a generalisation

of a convexity inequality; is this a hint? Note finally that there is something slightly peculiar about the fact that we can prove the conjecture at both ends of a spectrum; namely, when the degrees of the vertices of the tree are as equal as possible (a path) and when they are as unequal as possible (a star). Does this remark contain the seed of a result?

Note that amongst the cases where the probability is greater than its classical value, in at least some of them it is persistently greater, in the sense that, as T varies in natural families, we have

$$\lim_{n\to\infty} \frac{\log(P\{T \text{ on } n \text{ vertices }\})}{n} > \alpha.$$

For example, in the simple case when $k = 2$, $\mathbf{s} = (s, 1-s)$ with $s \neq 1/2$ and P has zeroes on the diagonal with all other entries being one, it is easy to check that for a path the above limit is

$$(\log(s) + \log(1-s))/2 > \log(1/2) > \log(2s(1-s)) = \log(\alpha)$$

where we have used the strict concavity of log and that $4s(1-s) < 1$ for $s \neq 1/2$. Similarily, it is easy to check that this property holds for stars as the number of vertices goes to infinity in this model. It is not clear how generally this will hold.

2.5 Some further insight into the probabilities of cycles

In this section we show how the Mulholland-Smith theorem gives insight into when cycles are more or less probable than classically.

Theorem 2.26 *Let Q be a non-negative symmetric k by k matrix satisfying $\mathbf{v}^T Q \mathbf{v} = \alpha$, where α is a constant and $\mathbf{v}$ a fixed non-zero non-negative vector of norm 1. Then the Perron-Frobenius eigenvalue of Q is at least α. There is equality if and only if $\mathbf{v}$ is an eigenvector of Q.*

Proof. We recall the standard fact of linear algebra that for a symmetric k by k matrix Q, we have, letting $\lambda_1 \geq \lambda_2 \geq ...\lambda_k$ be the eigenvalues of Q,

$$\max_{\mathbf{v}:\, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T Q \mathbf{v} = \lambda_1$$

where λ_1 is the largest eigenvalue; there is equality if and only if $\mathbf{v}$ is an eigenvector of Q. Indeed writing $\mathbf{v} = \sum_{i=1}^{k} \mu_i \mathbf{e}_i$ where the $\mathbf{e}_i$ are an orthonormal

basis of eigenvectors of Q, so that $\sum_{i=1}^{k} \mu_i^2 = 1$, we have

$$\mathbf{v}^T Q \mathbf{v} = \sum_{i}^{k} \mu_i^2 \lambda_i \leq \lambda_1$$

since the μ_i^2 are all non-negative. To get equality, we must have that the sum of the μ_i^2 for those λ_i which are equal to λ_1 must be one; thus $\mathbf{v}$ is indeed an eigenvector of Q. Thus we have, since $\mathbf{v}$ is of norm 1,

$$\alpha = \mathbf{v}^T Q \mathbf{v} \leq \max_{\mathbf{w}: \mathbf{w}^T \mathbf{w} = 1} \mathbf{w}^T Q \mathbf{w} = \lambda_1$$

with equality if and only if $\mathbf{v}$ is an eigenvector of Q as required. •

Corollary 2.27 *If r is even, the probability of an r-cycle is at least as large as classically in any RRC model. Also, if there is an even $r \geq 4$ for which the probability of an r-cycle is the same as classically, the model is trivial.*

Proof. The first sentence is immediate from Theorem 2.10 and Theorem 2.26. For the second, we note that if the probability of an r-cycle is α^r, then as the greatest term in Theorem 2.10 is $\lambda^r \geq \alpha^r$ and all the other terms are non-negative, we must have $\lambda = \alpha$ and all the other eigenvalues are zero; thus the model is CID and so trivial by Theorem 2.16. •

The analogous statement for cycles of odd length is false; there are models which are TID, give 3-cycles the same probability as classically, and are not trivial. For example, with three equiprobable colours, if P satisfies $p_{11} = p_{23} = p_{32} = \beta \in [0,1]$ and all other entries are zero; then the model is non-trivial, the probability of a 3-cycle is $(\beta/3)^3 = \alpha^3$ since all vertices must be colour 1 for the triangle to have a chance of forming, and the model is TID since the colours are equiprobable and all row sums of P are equal. In fact more generally, the probability of any cycle of odd length is the same as classically by the same argument. Thus, whilst it is sufficient to check the TID condition on the simplest tree imaginable (the path $1-2-3$), it is not enough to check the CID condition on the simplest cycle imaginable. We can generalise this example as follows.

Theorem 2.28 *Suppose $\Gamma(n,k,\mathbf{s},P)$ is a family of RRC models where P does not depend on n and the probability of any cycle of odd length is the same as classically. Then there is at least one eigenvalue of Q equal to α, and those of the other $k-1$ eigenvalues of Q which are non-zero occur in pairs of equal modulus and opposite sign. In particular, the rank of Q is odd.*

Proof. We again use the formula of Theorem 2.10. If the top eigenvalue is greater than α it must have a matching eigenvalue of equal modulus and opposite sign, as if all other eigenvalues are of smaller modulus it would dominate the formula for large enough r, making the probability of the cycle greater than classically; on the other hand, this eigenvalue cannot have modulus greater than the top eigenvalue as then the probability of long enough cycles would be negative. Thus we pair off the top eigenvalue, and then repeat the argument on the next largest eigenvalue greater than α (if any; this eigenvalue may equal the top eigenvalue since Q is imprimitive). Thus all unpaired eigenvalues are at most α. Thus we must have an eigenvalue α, as otherwise all the eigenvalues would be less than α and for large enough r their sum would not be as large as α^r. For each further eigenvalue equal to α there must be a matching one of opposite sign, to keep the total probability of the cycle equal to α^r. Now consider the eigenvalues less than α; as the sum of their r-th powers is zero for all odd r by the above remarks, we see, considering the largest of these eigenvalue(s) first, that each must have a matching eigenvalue of opposite sign; we then apply the same argument to the next largest eigenvalue, and so on. This gives the first statement of the result, and the second is an immediate corollary. •

This allows us to deduce that the general Q with this property is a simple generalisation of the one in the example above.

Theorem 2.29 *Suppose a family of RRC models, with P not depending on n, has the same probability of all cycles of odd length as classically. Then there is an orthonormal basis $\mathbf{e}_i$, $1 \leq i \leq k$ consisting of $\mathbf{e}_1$ an eigenvector of Q with eigenvalue α, the eigenvectors with eigenvalue 0, and the remaining basis elements can be partitioned into pairs on which Q acts by a 2 by 2 matrix with zeroes on the diagonal and a positive constant β off the diagonal.*

Proof. Let $\mathbf{w}$ and $\mathbf{u}$ be orthonormal eigenvectors of Q whose corresponding eigenvalues are $\beta > 0$ and $-\beta$. Then

$$Q(\mathbf{w} + \mathbf{u}) = \beta(\mathbf{w} - \mathbf{u}) \text{ and } Q(\mathbf{w} - \mathbf{u}) = \beta(\mathbf{w} + \mathbf{u}).$$

In other words, Q acts on the space generated by $\mathbf{w} + \mathbf{u}$ and $\mathbf{w} - \mathbf{u}$ by interchanging the two vectors and dilating them, so that on this 2-dimensional subspace it acts by the 2 by 2 matrix with zeroes on the diagonal and β elsewhere. Now by Theorem 2.28 all the eigenvectors other than one with eigenvalue α and those with eigenvalue zero come in pairs like this, and so

with respect to the basis obtained from the earlier one by replacing each such pair of eigenvectors $\mathbf{w}$ and $\mathbf{u}$ with $(\mathbf{w}+\mathbf{u})/\sqrt{2}$ and $(\mathbf{w}-\mathbf{u})/\sqrt{2}$, which it is easy to check still form an orthonormal basis, we get the claim. •

We now draw some more consequences of Theorem 2.26.

Corollary 2.30 *If the model is not TID, the probability of any cycle of large enough length is greater than classically. In particular, in $G_{p,q,r}$ with $r \neq p$, the probability of any long enough cycle is greater than classically (even when $q > \max\{p,r\}$).*

Proof. Combining Theorem 2.26 with Lemma 2.23, we see that $\lambda > \alpha$, and then the result follows considering large enough values of r the cycle length by using Theorem 2.10. •

We emphasise that Corollary 2.30 is a result about long enough cycles, not all cycles. We use the following lemma;

Lemma 2.31

$$P\{1-2-3-1 \text{ in } {}_sG_{p,q,r}\} - P\{1-2-3-1 \text{ in } G_\alpha\}$$

$$= s^3p^3+3s^2(1-s)pq^2+3(1-s)^2srq^2+(1-s)^3r^3-(s^2p+2s(1-s)q+(1-s)^2r)^3.$$

Proof. This expression is obtained from a simple calculation, conditioning on the eight possible colourings of the three vertices. •

Lemma 2.32 *There exist s,p,q and r for which ${}_sG_{p,q,r}$ has $\mathbf{v}$ not an eigenvector of Q, and a cycle whose probability is less than classically.*

Proof. For the first part, Theorem 2.8 suggests that we should take (say) $p = 1/2$, $r = 11/20$ and $q > \max\{r,p\}$ as then it seems likely that, as p and r are close, the probability will still be less than classically for suitably large q. And indeed taking $s = 1/2$, $p = 1/2$ $r = 11/20$ and $q = 7/10$ in the previous lemma, we clearly have that $\mathbf{v}$ is not an eigenvector and the expression in Lemma 2.31 evaluates to about $-0.000424 < 0$ as required. •

Exactly how long a cycle of odd length has to be to guarantee that it is at least as likely to arise as classically under the conditions of Theorem 2.30 will depend heavily on the nature of Q and $\mathbf{v}$ and seems difficult to say much about in general.

The case of $G_{p,q}$ with both p and q positive so that Q is primitive but $q > p$ so that the probability of cycles of any odd length is less than classically

by Corollary 2.8 shows that the restriction in Theorem 2.30 that $\mathbf{v}$ should not be an eigenvector of Q is genuinely necessary and not just a restriction of our method. Needless to say, having all odd length cycles no more likely than classically is quite a demanding condition;

Corollary 2.33 *Suppose the probability of an r-cycle, for all odd r, is less than or equal to its classical value. Then the largest eigenvalue for which there is not an eigenvalue of equal modulus and opposite sign is at most α. If there is equality the eigenvalue of maximum modulus amongst those remaining must be non-positive.*

Proof. This is very similar to Theorem 2.10 or Theorem 2.28. Again we must have that any eigenvalue greater than α pairs off with one of opposite sign and equal modulus. Thus if μ is the largest unpaired eigenvalue, $\mu \leq \alpha$; again there is no unpaired negative eigenvalue of modulus greater than μ as otherwise the probability of the cycle would be negative, so μ has the largest modulus of the unpaired eigenvalues. If $\mu = \alpha$, then the sum of the r-th powers of the remaining eigenvalues is non-positive for all r odd, and so the largest in modulus of them must be non-positive. •

The other question which demands attention is how much larger than classically λ can be. Theorem 2.9, part 5, gives an upper bound, namely the largest row sum of the matrix Q (the lower bound in that theorem is of course superseded by Theorem 2.26); it is not clear if one can significantly improve this upper bound in our special circumstances.

3 Joint probabilities of subgraphs and the role of the FKG inequalities

3.1 The main theorem on joint probabilities

In the previous chapter we discussed how the probabilities of certain simple kinds of subgraphs compare with their classical values, obtaining results which held or were conjectured to hold, in wide generality. In this section, we compare the joint probability of **any** two potential subgraphs arising with the product of their individual probabilities (we shall often abbreviate potential subgraph by subgraph if there is no danger of confusion); this is a more general question, since the subgraph is now arbitrary, but our results will work less generally. Our main result (Theorem 3.1) is that, in $G_{p,q}$, if $p > q$, the probability that two subgraphs both arise is at least as large as the product of their individual probabilities. Much of the rest of the chapter consist of glosses on, extensions of and complements to Theorem 3.1, and counterexamples to putative extensions. We also show that the so-called FKG inequality, which experience of the classical model might suggest is an appropriate tool for proving Theorem 3.1, will not give the result, and show that various related notions of association do not apply either. Finally, we consider what can be said for 3-cycles by doing exact calculations.

We shall write, if C is a graph on $\{1, 2, ...n\}$ (recall again that graphs are labelled unless explicitly stated otherwise) $P\{C\}$ for the probability that all the edges in C arise (again, we do not ask about the other edges). We work in $G_{p,q}$ until section 3.5, and will be comparing, for two subgraphs C_1 and C_2, $P_{p,q}\{C_1 \cap C_2\}$ with $P_{p,q}\{C_1\}P_{p,q}\{C_2\}$; here $P\{C_1 \cap C_2\}$ means the probability that the graph contains both the subgraphs C_1 and C_2. In particular, note that when C, C_1, C_2 and so forth are used in this chapter, they are not arbitrary events; they are events of the form all the edges in some set arise.

Of course, if the two subgraphs have neither edges nor vertices in common, Lemma 2.1 says that whether or not they arise is independent. This is also true in $G_{p,q}$ if they have exactly one vertex v (and so no edges) in common, since

$$P\{C_1 \cap C_2\} = \frac{P\{C_1 \cap C_2 \mid c(v) = \text{red}\}}{2} + \frac{P\{C_1 \cap C_2 \mid c(v) = \text{blue}\}}{2};$$

by symmetry (as the colours are equiprobable) the probability of each sub-

graph conditional on $c(v)$ is independent of the particular colour $c(v)$; hence

$$P\{C_1 \cap C_2\} = P\{C_1 \cap C_2 \mid c(v) = \text{red}\} = P\{C_1 \cap C_2 \mid c(v) = \text{blue}\}$$

so we write $P\{C_1 \cap C_2 \mid c(v)\}$ for their common value, and note that in the same way we can write $P\{C_i\} = P\{C_i \mid c(v)\}$. Then, using the fact that C_1 and C_2 are independent conditional on $c(v)$, we have

$$P\{C_1 \cap C_2\} = P\{C_1 \cap C_2 \mid c(v)\} = P\{C_1 \mid c(v)\}P\{C_2 \mid c(v)\} = P\{C_1\}P\{C_2\}.$$

However, the existence of two subgraphs with only vertices in common is not in general independent in $G_{p,q}$. A simple example is when C_1 is the path $1-2-3$ and C_2 the edge $1-3$; then by Theorem 2.4 $P\{C_1\}P\{C_2\} = \alpha^3$ but by Theorem 2.6 $P\{C_1 \cap C_2\} = ((p+q)/2)^3 + ((p-q)/2)^3$. More generally if we break any r-cycle up into two paths, Theorems 2.6 and 2.4 tell us that (provided $p > q$)

$$P\{C_1 \cap C_2\} = (\frac{p+q}{2})^r + (\frac{p-q}{2})^r \geq (\frac{p+q}{2})^r = P\{C_1\}P\{C_2\}.$$

Theorem 3.1 will show that this is an example of a general phenomenon. In the proof, we shall mark two inequalities (&) and (*) to which we shall make repeated reference later in the chapter. We shall also spell some points in the proof out rather carefully to facilitate drawing corollaries later.

Theorem 3.1 *Let C_1 and C_2 be two potential subgraphs on n vertices. Then, if $p > q$,*

$$P_{p,q}\{C_1 \cap C_2\} \geq P_{p,q}\{C_1\}P_{p,q}\{C_2\}.$$

Proof. First of all, we assume for convenience that $q \neq 0$, dealing with the special case $q = 0$ separately later.

We assume to begin with that C_1 and C_2 have no edges in common, so that they are **edge-disjoint**. Define S_i to be the number of edges of C_i whose two end vertices are the same colour (call such edges **non-switches**) and let n_i be the number of edges in C_i. Then, as C_1 and C_2 are edge-disjoint

$$P\{C_1 \cap C_2\} = \mathbf{E}\left(p^{S_1+S_2}q^{n_1+n_2-S_1-S_2}\right) = q^{n_1+n_2}\mathbf{E}\left((\frac{p}{q})^{S_1+S_2}\right)$$

$$= q^{n_1+n_2}\mathbf{E}\left(e^{\theta(S_1+S_2)}\right),$$

where $\theta = \log(p/q)$ which exists and is positive since $p > q > 0$. Similarily

$$P\{C_1\} = q^{n_1}\mathbf{E}\left(e^{\theta S_1}\right) \text{ and } P\{C_2\} = q^{n_2}\mathbf{E}\left(e^{\theta S_2^*}\right)$$

where S_2^* is the number of non-switches in C_2^*, a copy of the colouring of C_2 which has neither vertices nor edges in common with C_1, so that S_1 and S_2^* are independent. Then we have

$$P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\} = q^{n_1+n_2}\mathbf{E}\left(e^{\theta(S_1+S_2)}\right) - q^{n_1}\mathbf{E}\left(e^{\theta(S_1)}\right) q^{n_2}\mathbf{E}\left(e^{\theta(S_2^*)}\right)$$

$$= q^{n_1+n_2}\left(\mathbf{E}\left(e^{\theta(S_1+S_2)}\right) - \mathbf{E}\left(e^{\theta(S_1+S_2^*)}\right)\right)$$

as S_1 and S_2^* are independent. Thus, as $q \neq 0$, it suffices to show

$$\mathbf{E}\left(e^{\theta(S_1+S_2)}\right) \geq \mathbf{E}\left(e^{\theta(S_1+S_2^*)}\right)$$

for which it is sufficient to prove that the two Taylor expansions satisfy

$$\sum_{r=0}^{\infty} \frac{\theta^r \mathbf{E}((S_1+S_2)^r)}{r!} \geq \sum_{r=0}^{\infty} \frac{\theta^r \mathbf{E}((S_1+S_2^*)^r)}{r!} \quad (\&).$$

Since as we noted above $\theta > 0$, this will follow if we can show that

$$\mathbf{E}\left((S_1+S_2)^r\right) \geq \mathbf{E}\left((S_1+S_2^*)^r\right) \ \forall r.$$

Towards this, number the edges of C_1 and C_2 in any way, giving C_2^* the numbering induced by that on C_2, and let $S_1 = \sum I_i$, where I_i is an indicator of whether the i-th edge of C_1 is a non-switch, $S_2 = \sum J_j$, where J_j is an indicator of whether the j-th edge of C_2 is a non-switch and $S_2^* = \sum J_j^*$ where J_j^* is an indicator of whether the j-th edge of C_2^* is a non-switch. Then, considering the binomial expansions of $\mathbf{E}((S_1+S_2)^r)$ and $\mathbf{E}((S_1+S_2^*)^r)$ for all values of r, we see it suffices to show that

$$\mathbf{E}\left(I_{a_1}I_{a_2}...I_{a_m}J_{b_1}...J_{b_l}\right) \geq \mathbf{E}\left(I_{a_1}I_{a_2}...I_{a_m}J_{b_1}^*...J_{b_l}^*\right) \quad (*)$$

for all choices of $a_1,...a_m$ and $b_1,...b_l$. In fact we can restrict to the case where $m \leq n_1$ and $l \leq n_2$ as otherwise some I, J or J^* occurs to a power higher than 1, but because the I_i, J_j and J_j^* are indicator variables this reduces to the case $m \leq n_1$ and $l \leq n_2$.

Note that, in each component of the graph whose edges are those involved in the expression $(*)$, a product of indicators is 1 if and only if all

the indicators are 1, i.e all the vertices involved are the same colour, and is 0 otherwise.

Suppose first $m + l < \min(g(C_1), g(C_2))$ where $g(C)$ is the girth of the graph C, that is the length of the shortest cycle, so that no cycle in C_1 and no cycle in C_2^* is present on the right-hand side of $(*)$. Then the right-hand side of the equation $(*)$ is equal to the probability that the product of indicators there is 1, which is the probability that all the Is and J^*s involved are equal to 1. As C_1 and C_2^* have neither edges nor vertices in common, all $m + l$ indicators are of different edges, and since there are no cycles, at each stage when we check whether or not the next edge is a non-switch, we know the colour of one of the vertices, and not that of the other; since we are in $G_{p,q}$ they are the same with probability $1/2$, so that the right-hand side is $(1/2)^{m+l}$ since $m + l$ is the number of edges being considered. If there are no cycles on the left-hand side of the equation $(*)$ we will get the same value there; however there can be cycles on the left-hand side since C_1 and C_2 interact with each other (for example if I_1 is the edge $1-2$, I_2 the edge $2-3$ and J_1 the edge $3-1$). Such cycles mean that the indicators are correlated; in the example, if I_1=1 and I_2=1, J_1 is automatically also 1. However by stripping away enough of the indicators (t of them, say) to reduce to the indicators of a maximal forest in the collection of edges being examined we regain independence of the remaining indicators, and the left-hand side of $(*)$ is $(1/2)^{m+l-t} > (1/2)^{m+l}$ as required.

Next we drop the condition $m + l < \min(g(C_1), g(C_2))$ so thus there may be some cycles in the two subgraphs C_1 and C_2 present. However any such cycle present on the right-hand side of $(*)$ will also be present on the left-hand side, so the correlation effect appears, making an equal contribution, on both sides; moreover, as in the preceding paragraph, there can be further correlation on the left-hand side of $(*)$ due to cycles forcing some of the indicators to be 1, as illustrated with I_1, I_2 and J_1 above, so again the result holds.

Next we show that removing the edge-disjointness condition can only increase $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$. Since $P\{C_1\}$ and $P\{C_2\}$ will not change, it suffices to show $P\{C_1 \cap C_2\}$ increases. To see this, note that the formula for $P\{C_1 \cap C_2\}$ will now double count some of the edges so giving a larger power of p or q or both than is in fact appropriate, thus underestimating the true value of $P\{C_1 \cap C_2$, and $P\{C_1\}$ and $P\{C_2\}$ are of course unchanged. Thus if suffices to show that $(*)$ still holds; and the right-hand side of $(*)$ is not affected by the change, as C_1 and C_2^* are disjoint by construction;

however, on the left-hand side, additional correlation structure may increase the expectation.

Thus it only remains to deal with the case $q = 0$. Here the subgraphs arise if and only if all the vertices in each component are the same colour. But again, the correlation structure, through cycles, can force some vertices to be the same colour when both graphs are present, when they are not forced to be the same colour in the independent copies of the two subgraphs, and so we get the required inequality. Alternatively we could just argue by the continuity of the probability of C as a function of q. •

Corollary 3.2 *Let C_1, $C_2 ... C_k$ be k subgraphs. Then in $G_{p,q}$ with $p > q$*

$$P\{C_1 \cap C_2 \cap ... \cap C_k\} \geq P\{C_1\}P\{C_2\}...P\{C_k\}$$

Also if C_1, C_2,..C_r are some of these subgraphs

$$P\{C_1 \cap C_2 \cap ... \cap C_r \mid C_{r+1} \cap ... \cap C_k\} \geq P\{C_1\}P\{C_2\}...P\{C_r\}$$

Proof. The proof of the first claim is by induction on k. The case $k = 2$ is Theorem 3.1, and as several subgraphs together are a subgraph, we have

$$\begin{aligned} P\{C_1 \cap C_2 \cap ... \cap C_k\} &= P\{(C_1 \cap C_2 \cap ... \cap C_{k-1}) \cap C_k\} \\ &\geq P\{C_1 \cap C_2 \cap ... \cap C_{k-1}\}P\{C_k\} \text{ by Theorem 3.1} \\ &\geq P\{C_1\}P\{C_2\}...P\{C_k\} \text{ by the induction hypothesis.} \end{aligned}$$

For the second claim, it is sufficient to show that

$$P\{C_1 \cap C_2 \cap ... \cap C_k\} \geq P\{C_1\}P\{C_2\}...P\{C_r\}P\{C_{r+1} \cap ... \cap C_k\}.$$

for which, by the first part of this corollary, it suffices to prove

$$P\{C_1 \cap C_2 \cap ... \cap C_k\} \geq P\{C_1 \cap C_2 ... \cap C_r\}P\{C_{r+1} \cap ... \cap C_k\}.$$

However this is a consequence of Theorem 3.1 on noting that the graph consisting of all edges in one or more of the graphs C_1, $C_2 ... C_r$ and the graph consisting of all edges in one or more of the graphs C_{r+1}, $...C_k$ are both themselves subgraphs. •

Theorem 3.3 *Let S_i be the number of edges of C_i which are non-switches as above. Then for non-negative integers r and s we have*

$$\mathbf{E}(S_1^r S_2^s) \geq \mathbf{E}(S_1^r)\,\mathbf{E}(S_2^s).$$

Proof. This follows from expanding out the two expressions in the statement of the theorem as sums of indicator functions, and making repeated use of inequality $(*)$ in the proof of Theorem 3.1. •

Corollary 3.4 *In $G_{p,q}$ with $p > q$ all moments $\mathbf{E}(N^r)$ of N, the number of cycles in the graph, are at least as large as the corresponding ones in the corresponding classical model. The same is true for N_k the number of k-cycles for any $k \geq 3$.*

Proof. N is a sum of the indicators of whether each possible cycle arises. Now expand out N^r to get a sum of products of such indicators. Each such product is 1 or 0, being 1 only if all the cycles are present. But by Theorem 3.1 the probability that they are all present is at least as large as classically, and the result follows. The proof for N_k is identical. •

An argument similar to the argument in the proof of Theorem 3.1 will be found in Theorem 4.12 below. Theorem 3.1 is closely related to a result of Harris [Ha] in the theory of percolation on the two-dimensional integer lattice where bonds (edges) are open (arise) with probability p and closed (fail to arise) with probability $1-p$, independently of each other. However the method of proof is quite different. We will show in section 3.8 that unlike Harris's result, ours cannot be put in the context of the FKG inequality.

Note that the argument does depend on the fact that $p > q$ in that it uses $\theta = \log(p/q) > 0$ to reduce proving the inequality between Taylor series to an inequality between Taylor coefficients. If $p < q$, so that $\theta < 0$ this will not work; for example, if C_1 is the path $1-2-3$ and C_2 the edge $1-3$,

$$P\{C_1 \cap C_2\} = (\frac{p+q}{2})^3 + (\frac{p-q}{2})^3 < (\frac{p+q}{2})^3 = P\{C_1\}P\{C_2\}.$$

On the other hand, it is not true that $P\{C_1 \cap C_2\} < P\{C_1\}P\{C_2\}$ always holds when $q > p$ either, because if r is even the probability of an r-cycle is greater than classically by Theorem 2.6. We discuss the case $p < q$ in detail in section 3.3.

If $p = q$ of course $P\{C_1 \cap C_2\} = P\{C_1\}P\{C_2\}$ if C_1 and C_2 are edge disjoint, and $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ otherwise, with equality only if

$p = q = 0$, as otherwise α^{2n-t}, where t is the number of common edges, exceeds α^{2n}.

Some special cases of Theorem 3.1 can be proven more directly; for example, if two edge-disjoint cycles have r common vertices which can be numbered so that there is exactly one arc of C_1 and one arc of C_2 between m and $m+1$ for each m (including of course arcs between r and 1 to close the cycles) then a proof very similar to Theorem 2.6 gives the result, though of course not all pairs of cycles are of this form.

Finally note that the fact that the colours are equiprobable is not needed in Theorem 3.1. Indeed the probabilities of the colours only come into the argument at the point where we have to prove inequality $(*)$, namely that

$$\mathbf{E}\left(I_{a_1}I_{a_2}...I_{a_m}J_{b_1}...J_{b_l}\right) \geq \mathbf{E}\left(I_{a_1}I_{a_2}...I_{a_m}J^*_{b_1}...J^*_{b_l}\right)$$

and remark that the right-hand side is $(1/2)^{(m+l)}$ in the case when there are no cycles present on the right-hand side (with obvious modifications for the case where there are cycles on the right-hand side). However, all that matters for this argument is that (assuming that the $m+l$ indicators on the right-hand side form a connected component; if not, just argue componentwise) the right-hand side is $s^{m+l} + (1-s)^{m+l}$ where s is the probability that two vertices are both red, whereas the left-hand side will be $s^{m+l-t} + (1-s)^{m+l-t}$ for some $t \geq 0$, and thus will be at least as large as the right-hand side. However we shall see later, when we try to extend Theorem 3.1 to ${}_sG_{p,q,r}$, that the probabilities of the various colours do become important, at least to our method of proof. The assumption that different vertices are coloured independently is important, as otherwise it would be much harder to understand the variables S_1, S_2, etc.

3.2 Detailed comparison of the joint and individual probabilities

We now make a more detailed comparison of the two Taylor series in Theorem 3.1 (see equation (&)). We make the following definition.

Definition 3.1 *The* **newgirth** *of two (labelled!) subgraphs C_1 and C_2 is the length of the shortest cycle in the graph whose edges are those of C_1 and those of C_2 which is not in either C_1 or C_2. If no new cycle is introduced, we say the newgirth is infinity.*

Theorem 3.5 *Suppose the two subgraphs C_1 and C_2 are edge-disjoint. Then the two Taylor expansions in inequality (&) have coefficients which agree up to and including the term in θ^{g-1}, where g is the newgirth of C_1 and C_2. (If the newgirth is infinity, the claim is that the two series are identical).*

Proof. The terms of the two Taylor expansions in θ^k involve (writing the S_i as sums of indicators as in the proof of Theorem 3.1) only products of at most k indicators. But the k indicators cannot be those of a new cycle caused by putting the two subgraphs together unless $k \geq g$, and having a new cycle is the only way we get the Taylor series on the left-hand side of (&) to be different from the Taylor series on the right-hand side of (&) as we saw in the proof of Theorem 3.1. •

Corollary 3.6 *S_1 and S_2 are uncorrelated if C_1 and C_2 are edge disjoint.*

Proof. S_1S_2 is a sum of products of pairs of indicators, and a pair of edges cannot define a cycle, so by the previous result

$$\mathbf{E}(S_1S_2) = \mathbf{E}(S_1S_2^*) = \mathbf{E}(S_1)\mathbf{E}(S_2^*) = \mathbf{E}(S_1)\mathbf{E}(S_2). \bullet$$

However the two variables are not independent; for example, suppose C_1 and C_2 are both cycles so of length n (so they pass through each of the n vertices of the graph); then if S_1 is n, all the vertices are the same colour so S_2 must also be n.

The difference between the first two distinct coefficients in the two power series in Theorem 3.5 can be quite large. For example, if C_1 is a $(2n+1)$-cycle $1-2-3-....-(2n+1)-1$ and C_2 is also a $(2n+1)$-cycle $1-3-5-.....(2n+1)-2-4-...(2n)-1$, we see that there are a full $(2n+1)$ triangles introduced; this is the maximum number possible since the edge with the lowest number can be chosen in $(2n+1)$ ways and everything else is then forced; in this case thus the difference between the two Taylor coefficients is $(2n+1)(1/4-1/8) = (2n+1)/8$ since the probability that three edges on the right-hand side of (&) are all non-switches is $1/8$ but on the left-hand side if two of them are non-switches so is the third, so the probability is $1/4$. Presumably taking two long cycles, we can get arbitrarily large newgirth but still have the difference between the coefficients growing linearly with n.

3.3 Comparison of joint and individual probabilities when $q > p$

We now consider what can be said in the harder case when $q > p$.

Theorem 3.7 *Suppose $q > p$, and that C_1 and C_2 are edge-disjoint. Then there is a neighbourhood $q \in (\alpha, \alpha+\epsilon)$ where $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\} > 0$ or is < 0 according as the newgirth is even or odd.*

Proof. By the argument in Theorem 3.5, the coefficients of the two Taylor series agree for powers of θ less than g, where g is the newgirth of the two subgraphs, so we concentrate on the coefficient of θ^g. By the argument of Theorem 3.1, each product of g indicators on the right-hand side of equation $(*)$ in the proof of Theorem 3.1 is at least as large as the corresponding product on the left-hand side, and since we have taken the term in θ^g, there is at least one collection of indicators for which the left-hand side is actually greater than the right-hand side. As $\theta < 0$, $\theta^g > 0$ if g is even and is < 0 otherwise; thus, taking θ sufficiently close to 0 we see that there is a region $q \in (\alpha, \alpha + \epsilon)$ (where ϵ will depend on p, q, C_1 and C_2) in which we have $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\} > 0$ if g is even and < 0 if g is odd. •

Note that this, coupled with Theorem 3.1 for $p > q$ shows that for edge-disjoint subgraphs with even newgirth, $p = \alpha$ is a local minimum of the expression $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$.

This shows that if the newgirth is odd $P\{C_1 \cap C_2\} < P\{C_1\}P\{C_2\}$ for $q \in (\alpha, \alpha + \epsilon)$; however, whilst the inequality sometimes then holds for **all** $q > \alpha$ (as putting two paths together to make a cycle of odd length in $G_{p,q}$ shows, using Theorems 2.4 and 2.6) it does not always do so. We will show this by considering the case of two cycles C_1 and C_2 with three vertices and no edges in common. This is the simplest case of edge-disjoint cycles where there is any chance of a counterexample, as if C_1 and C_2 have at most one vertex in common we saw before Theorem 3.1 that they are independent, and if the two cycles intersect in exactly two vertices, with no edges in common, the following result implies there is no change of sign in the interval $(\alpha, 2\alpha)$.

Theorem 3.8 *Suppose the cycles C_1 and C_2 have exactly two vertices a and b (and no edges) in common. Then $P\{C_1 \cap C_2\} < P\{C_1\}P\{C_2\}$ if and only if $p < q$ and the following statement holds for the two paths making up one of the two cycles, and does not hold for the two paths making up the other cycle;*

Both the paths between a and b in the cycle are of odd length, or the longer of these two paths is of even length and the shorter of odd length.

Proof. The case $p = q = 0$ is trivial, so we assume at least one of p and q is strictly positive for the rest of the proof. C_1 consists of two arcs between the two common vertices; let them be P_1 of length i and P_2 of length j, and similarily let C_2 consist of P_3 of length k and P_4 of length l. To keep notation under control, we will write α for $(p+q)/2$ and β for $(p-q)/2$; then, conditioning on the possible colourings of the two common vertices, and using Corollary 2.7 to find the probabilities of the various paths, we see, using computer simplification, that $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$ is

$$\left(\alpha^k\beta^l + \beta^k\alpha^l\right)\left(\alpha^i\beta^j + \beta^i\alpha^j\right).$$

which is < 0 if and only if exactly one of the two factors is. Clearly neither factor can be negative if $\beta > 0$; if $\beta < 0$ then, dividing by $\left(\frac{p+q}{2}\right)^{k+l}$ as $p+q > 0$, the first factor is negative if and only if

$$\left(\frac{p-q}{p+q}\right)^k + \left(\frac{p-q}{p+q}\right)^l < 0.$$

which in turn holds if and only if k and l are both odd **or** the smaller of them is odd and the larger even. The statement of the theorem is now clear. •

We can now proceed to the promised counterexample. Take two edge-disjoint cycles which have three vertices a, b and c in common. Let C_1 consist of the paths P_1, P_2 and P_3 and C_2 of the paths P_4, P_5 and P_6, with P_1 and P_4 going from a to b, P_2 and P_5 from b to c and P_3 and P_6 from c to a; let the lengths of P_1,...P_6 be i, j, k, r, s and t respectively. Conditioning on the eight possible colourings of a, b and c, and again using Corollary 2.7 to find the probabilities of a path conditional on the colour of its end vertices, computer simplification gives

$$\begin{aligned}
&P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\} \\
&= \beta^{i+j+t}\alpha^{k+r+s} + \beta^{i+j+r+s}\alpha^{k+t} + \beta^{i+k+s}\alpha^{j+r+t} + \beta^{i+k+r+t}\alpha^{j+s} \\
&+\alpha^{i+j+t}\beta^{k+r+s} + \alpha^{i+k+r+t}\beta^{j+s} + \alpha^{i+k+s}\beta^{j+r+t} + \alpha^{i+r}\beta^{j+k+s+t} \\
&+\alpha^{i+s+t}\beta^{j+k+r} + \alpha^{i+j+r+s}\beta^{k+t} + \beta^{i+s+t}\alpha^{j+k+r} + \beta^{i+r}\alpha^{j+k+s+t}
\end{aligned}$$

The result then follows considering the case when $i = 2$, $j = 2$, $k = 1$, $r = 1$, $s = 2$, $t = 3$, when the formula becomes (reverting to p and q now)

$$\frac{3p^{11}}{512} - \frac{p^9q^2}{512} - \frac{9p^7q^4}{256} + \frac{15p^5q^6}{256} - \frac{17p^3q^8}{512} + \frac{3pq^{10}}{512}$$

and putting $q = zp$ and removing the factor p^{11}, the resulting polynomial in z is

$$\frac{3}{512} - \frac{z^2}{512} - \frac{9z^4}{256} + \frac{15z^6}{256} - \frac{17z^8}{512} + \frac{3z^{10}}{512}$$

which has roots at (amongst other places) $z = \sqrt{3}$, which is about 1.73; at $z = 1.7$ the polynomial is negative but at $z = 1.8$ it is positive. It may well be possible to get, with more complicated C_1 and C_2, cases where the set of q (for fixed α) such that $P_{p,q}\{C_1 \cap C_2\} \geq P_{p,q}\{C_1\}P_{p,q}\{C_2\}$ has several connected components.

3.4 Detailed comparison for non-edge-disjoint subgraphs

We saw in Theorem 3.1 that when the cycles are not edge-disjoint, we have (for $p > q$) $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$. It is arguably more appropriate in this case to take account of the fact that the cycles are not edge-disjoint, and so that edges will be counted twice, and see if any more detailed result can be proved. One naive idea along these lines uses the following definition;

Definition 3.2 *The* **intersection number** $i(C_1, C_2)$ *of two subgraphs* C_1 *and* C_2 *is the number of common edges of the two subgraphs.*

We might then consider just how small we can take κ, independent of the choice of C_i but depending on p and q, such that

$$\kappa^{i(C_1,C_2)}P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}.$$

$\kappa = 1$ is clearly possible by Theorem 3.1; in G_α we can take $\kappa = \alpha$ as is easy to see. One result to improve on $\kappa = 1$ is easy;

Theorem 3.9 *Let* C_1 *and* C_2 *be subgraphs in* $G_{p,q}$*, with* $p > q$*. Then if* $\kappa = p$*, we have*

$$\kappa^{i(C_1,C_2)}P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$$

Proof. Recall that in the last paragraph of the proof of Theorem 3.1 we saw that, because the two subgraphs have edges in common,

$$P\{C_1 \cap C_2\} \geq \mathbf{E}\left(p^{S_1+S_2} q^{n_1+n_2-S_1-S_2}\right)$$

since the right-hand side counts the common edges of the two subgraphs twice. Now for each of these t edges, we get an extra factor q or p on the right-hand side of the inequality. Since $p > q$ the extra factor is at most p in each case, so in fact if we put a similar factor on the left-hand side we do not invalidate the inequality, and this gives the required result. •

One might hope to be able to get a κ less than p, since (loosely speaking) about half the edges to be double counted are same-different. One obvious guess would be that $\kappa = \alpha$ is possible, but this is not true. Indeed if the two subgraphs have an r-cycle (for some r) amongst their common edges, as the probability of that cycle is greater than its classical value, a compensation factor of α is likely not to be enough; this gives the simple counterexample of taking C_1 and C_2 to be the same triangle; then $i(C_1, C_2) = 3$ but

$$\alpha^3 P\{C_1 \cap C_2\} = \alpha^3\left(\alpha^3 + \left(\frac{p-q}{2}\right)^3\right) \leq \left(\alpha^3 + \left(\frac{p-q}{2}\right)^3\right)^2 = P\{C_1\}P\{C_2\}.$$

This still leaves the possibility that we can take $\kappa = \alpha$ when the two subgraphs have no cycle amongst their m common edges. Attempts to prove this along the lines of Theorem 3.1 seem not to work, since we now need a further random variable S_3 the number of non-switches which are in both subgraphs; then of course S_3^*, defined in the obvious way, will be zero, and so we need to show that, with $\theta = \log(p/q) > 0$ as before, that

$$\left(\frac{1+e^\theta}{2}\right)^m e^{\theta(S_1+S_2-S_3)} \geq e^{\theta(S_1+S_2^*)}$$

and this seems harder to handle than the previous expression; for example, we cannot just ignore the factor $(\frac{1+e^\theta}{2})^m \geq 1$ since if we could, that would be saying that we could take $\kappa = q$ which we shall see in a minute is impossible, and comparison of each coefficient of the two Taylor series looks daunting. Nonetheless, brute force methods, similar to those used in section 3.3 to show that the set of q where the joint probability of subgraphs exceeds the product of their probabilities need not be connected, enable us to prove the claim in several cases when C_1 and C_2 are cycles with small numbers of vertices and

edges in common, namely the following;

1. exactly one path (and no other vertices) in common

2. exactly two non-incident edges in common

3. one path and one other vertex in common

4. one path and two other non-adjacent vertices in common.

However it is not clear how to proceed in general, and so we offer an

Open Problem. Is it true that if C_1 and C_2 are two subgraphs, and the graph of their m common edges is cycle-free, then for $p > q$

$$\alpha^m P_{p,q}\{C_1 \cap C_2\} \geq P_{p,q}\{C_1\}P_{p,q}\{C_2\}?$$

Note that we certainly cannot do better than $\kappa = \alpha$. Indeed if C_1 is the subgraph $1-2-3$ and C_2 the subgraph $2-3-4$, $P_{p,q}\{C_i\} = \alpha^2$ in both cases, but $P_{p,q}\{C_1 \cap C_2\} = \alpha^3$ so we must have $\kappa \geq \alpha$.

3.5 Several colours but only two edge probabilities.

We next consider what can be said when there are more than two colours involved. In the special case where there are k colours (not necessarily equiprobable) and all same-same probabilities are equal to p and all same-different probabilities are q, an argument entirely analogous to Theorem 3.1 shows that $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ for $p > q$; S_i is defined as before to be the number of edges in C_i both of whose ends are the same colour, and the argument runs through. More generally, given k colours and all edges arising with probability q except those between two vertices both of colour i $(1 \leq i \leq r)$, where $r \leq k$, then $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ for $p > q$ by imitating the proof of Theorem 3.1 with the modification that S_i is now the number of edges of C_i between two vertices which are both the same colour j, for some $1 \leq j \leq r$, and the various indicator variables undergo an obvious analogous change of meaning. When we get to proving the analogue of $(*)$ it is enough to note that the product of indicators on the right-hand side $(*)$ is 1 if and only if for each edge involved both ends have the same colour out of the set

$\{1, 2...r\}$, and that as before some of the J_j may be forced to be 1 in cycles by the correlation structure whilst the J_j^* are not forced to be 1.

One might try to develop these ideas by introducing, given an RRC model with k colours and only two edge probabilities p and q, a (probably looped) graph, the **structure graph**, on the k colours as vertices with an edge between vertices i and j if and only if edges between a vertex of colour i and one of colour j arise with probability p, and asking for which structure graphs we have that $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ for $p > q$. So for example the case just discussed is the special case where the structure graph consists of loops at some vertices and no other edges. If, for example, the structure graph is complete (where now we insist a complete graph must include a loop at each vertex, unless it has only one vertex, when it either can or cannot have such a loop), the result holds rather trivially; if the components of the structure graph are all complete graphs, the property holds, essentially just by identifying all the colours in each complete graph; this works because in a graph all of whose components are complete, if i is connected to j, i is adjacent to j. However it seems rather harder to take this circle of ideas much further. The simplest structure graph not covered by the above observations is one with two equiprobable colours 1 and 2 with the only edge being $1 - 2$ (and no loops); this is (in light disguise) a $G_{p,q}$ model with $q > p$ and by Theorem 3.5 we can have $P\{C_1 \cap C_2\} \leq P\{C_1\}P\{C_2\}$ in this case.

3.6 The situation in ${}_sG_{p,q,r}$.

In general with more than two parameters matters are more complex. We first note that we cannot prove a statement analogous to Theorem 3.1 in $G_{p,q,r}$ with $r > p > q > 0$ by the same method; for, letting S_1 be the number of red-red edges in C_1, T_1 the number of blue-blue edges in C_1, and defining S_2, T_2, S_2^* and T_2^* in the obvious analogous manner, the joint probability is again a two variable generating function evaluated at $\theta_1 = \log(p/q)$ and $\theta_2 = \log(r/q)$, which exist and are positive by the assumptions, we reduce as in Theorem 3.1 to showing that

$$\mathbf{E}\left(((\theta_1 S_1 + \theta_2 T_1) + (\theta_1 S_2 + \theta_2 T_2))^r\right)$$

$$\geq \mathbf{E}\left(((\theta_1 S_1 + \theta_2 T_1) + (\theta_1 S_2^* + \theta_2 T_2^*))^r\right).$$

Now let $S_1 = \sum_{i=1}^{n_1} I_i$ where I_i is an indicator of whether the i-th edge of C_1 (with respect to some numbering of the edges of C_1) is red-red,

$S_2 = \sum_{i=1}^{n_2} J_j$ where J_j is an indicator of whether the j-th edge of C_2 (in some numbering of the edges of C_2) is red-red,

$S_2^* = \sum_{i=1}^{n_2} J_j^*$ where J_j^* is an indicator of whether the j-th edge of C_2^* (in the numbering induced by that on C_2) is red-red

$T_1 = \sum_{l=1}^{n_1} L_l$ where L_l is an indicator of whether the l-th edge (in the same numbering as for S_1 above) of C_1 is blue-blue,

$T_2 = \sum_{m=1}^{n_2} M_m$ where M_m is an indicator of whether the m-th edge of C_2 (in the same numbering as for S_2 above) is blue-blue,

$T_2^* = \sum_{n=1}^{n_2} M_m^*$ where M_m^* an indicator if the n-th edge of C_2^* is blue-blue.

Then, to get the obvious analogue of the previous argument to work, we would have to prove that

$$\mathbf{E}(I_{a_1}...I_{a_i}J_{b_1}...J_{b_j}L_{c_1}...L_{c_l}M_{d_1}...M_{d_m}) \geq \mathbf{E}(I_{a_1}...I_{a_i}J^*_{b_1}...J^*_{b_j}L_{c_1}...L_{c_l}M^*_{d_1}...M^*_{d_m})$$

for all choices of the subscripts, but it is clear that by judicious choice of C_1 and C_2 and the edges we can make the left-hand side zero but the right-hand side positive; for example, if C_1 and C_2 are cycles which intersect in two vertices a and b only, where there is an edge of C_1 between a and b, and a path of length two in C_2 between them, in the case where the indicators under consideration are those of whether the edge of C_1 is blue-blue, and the two edges in C_2 are both red-red; then clearly the left-hand side is zero, but the right-hand side is $\frac{1}{8}$.

To get round this in ${}_sG_{p,q,r}$ where $r > p > q$, define S_1 to be the number of edges in C_1 which are non-switches, as before, and S_2 and S_2^* similarily, but now define T_1 to be the number of **blue-blue** edges in C_1 and T_2 and T_2^* to be the number of blue-blue edges in C_2 and C_2^* respectively. Then

$$P\{C_1 \cap C_2\} = \mathbf{E}\left(r^{T_1+T_2}p^{S_1+S_2-T_1-T_2}q^{n_1+n_2-S_1-S_2}\right)$$

$$= q^{n_1+n_2}\mathbf{E}\left(\frac{r}{p}\right)^{T_1+T_2}\left(\frac{p}{q}\right)^{S_1+S_2}.$$

r/p and p/q are > 1 by the assumptions, and it now seems reasonable that

powers of the T_i and the S_j are non-negatively correlated. However, as the actual statement of the theorem shows, some caution is required.

Theorem 3.10 *Let C_1 and C_2 be two fixed graphs on vertex set $\{1,2,...n\}$. Then in ${}_sG_{p,q,r}$, if $r > p > q$ and $s \leq 1/2$ we have $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$.*

Proof. As in Theorem 3.1, we treat $q = 0$ as a special case, so assume for the moment that $q \neq 0$. We have

$$P\{C_1 \cap C_2\} = q^{n_1+n_2}\mathbf{E}\left(\frac{r}{p}\right)^{T_1+T_2}\left(\frac{p}{q}\right)^{S_1+S_2} = q^{n_1+n_2}\mathbf{E}\left(e^{\theta_1(T_1+T_2)+\theta_2(S_1+S_2)}\right)$$

where $\theta_1 = \log(r/p)$ and $\theta_2 = \log(p/q)$ exist and are > 0 as $r > p > q > 0$. Similarily

$$P\{C_1\} = q^{n_1}\mathbf{E}\left(e^{\theta_1 T_1+\theta_2 S_1}\right) \text{ and } P\{C_2\} = q^{n_2}\mathbf{E}\left(e^{\theta_1 T_2^*+\theta_2 S_2^*}\right)$$

where S_2^* is the number of non-switches in C_2^*, a copy of the colouring of C_2 which has neither vertices nor edges in common with C_1 (so that S_1 and S_2^* are independent), and T_2^* is the number of blue-blue edges in C_2^*. Hence

$$P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$$

$$= q^{n_1+n_2}\mathbf{E}\left(e^{\theta_1(T_1+T_2)+\theta_2(S_1+S_2)}\right) - q^{n_1}\mathbf{E}\left(e^{\theta_1 T_1+\theta_2 S_1}\right)q^{n_2}\mathbf{E}\left(e^{\theta_1 T_2^*+\theta_2 S_2^*}\right)$$

$$= q^{n_1+n_2}\mathbf{E}\left(e^{\theta_1(T_1+T_2)+\theta_2(S_1+S_2)}\right) - q^{n_1+n_2}\mathbf{E}\left(e^{\theta_1(T_1+T_2^*)+\theta_2(S_1+S_2^*)}\right)$$

since S_1 and S_2^* are independent and T_1 and T_2^* are independent. Hence (again by $q \neq 0$) it is enough to show that

$$\mathbf{E}\left(e^{\theta_1(T_1+T_2)+\theta_2(S_1+S_2)}\right) \geq \mathbf{E}\left(e^{\theta_1(T_1+T_2^*)+\theta_2(S_1+S_2^*)}\right);$$

expanding both sides out as bivariate Taylor series, this would follow from

$$\sum_{r=0}^{\infty}\sum_{i=0}^{r}\frac{\theta_1^i\theta_2^{r-i}\mathbf{E}((T_1+T_2)^i(S_1+S_2)^{r-i}}{i!(r-i)!} \geq \sum_{r=0}^{\infty}\sum_{i=0}^{r}\frac{\theta_1^i\theta_2^{r-i}\mathbf{E}((T_1+T_2^*)^i(S_1+S_2^*)^{r-i})}{i!(r-i)!}.$$

As $\theta_1 > 0$ and $\theta_2 > 0$, this will follow if we can show that

$$\mathbf{E}\left((T_1+T_2)^m(S_1+S_2)^n\right) \geq \mathbf{E}\left((T_1+T_2^*)^m(S_1+S_2^*)^n\right) \forall m,n.$$

Towards this, number the edges of C_1 and C_2 in any way, giving C_2^* the numbering induced by that on C_2. Then let

$T_1 = \sum I_i$, where I_i is an indicator of whether the ith edge of C_1 is blue-blue

$S_1 = \sum J_j$ where J_j is an indicator of whether the jth edge of C_1 is a non-switch.

$T_2 = \sum L_l$, where L_l is an indicator of whether the lth edge of C_2 is blue-blue

$S_2 = \sum M_m$, where M_m is an indicator of whether the mth edge of C_2 is a non-switch.

$T_2^* = \sum L_l^*$, where L_l^* is an indicator of whether the lth edge of C_2^* is blue-blue

$S_2 = \sum M_m^*$ where M_m^* is an indicator of whether the mth edge of C_2^* is a non-switch.

Then, considering the binomial expansions of $\mathbf{E}((S_1+S_2)^s)$ and $\mathbf{E}((T_1+T_2)^r)$, etc, for all values of r, we see it is enough to show that

$$\mathbf{E}(I_{a_1}..I_{a_i}L_{b_1}..L_{b_l}J_{c_1}..J_{c_j}M_{d_1}..M_{d_m}) \geq \mathbf{E}(I_{a_1}..I_{a_i}L^*_{b_1}..L^*_{b_l}J_{c_1}..J_{c_j}M^*_{d_1}..M^*_{d_m}) \quad (@)$$

for all possible choices of the suffices. In fact we can restrict to the case where no indicator occurs to a power higher than the first, since the variables involved are indicators.

As before, each expectation of a product of indicators is 1 if and only if all the indicators are 1. This happens if all the vertices in each component of the graph whose edges are indicated by the random variables involved are the same colour when the indicators involved are all Js, Ms or M^*s, and if all the vertices are blue if all the indicators involved are Is Ls or L^*s. Note also that if a component of the graph formed by the edges indexed by the Ms and Js has any vertex (or edge) in common with any component of the graph formed by the edges indexed by the Is and Ls then for both products of indicators to be 1 every vertex of both components must be blue. Hence, since $s \leq 1/2$, we see that the joint probability of the two events A, that all the edges in some component of the graph indexed by the Is and Ls are 1,

and the event B, that all the edges in some component of the graph indexed by the Ms and Js are 1, is at least as large as the product of the probability of A and the probability of B. For if the edges involved in the events A and B have neither edges nor vertices in common, the result follows from Lemma 2.1; otherwise, since they have common vertices, all the vertices in both components are forced to be blue, and if k is the number of vertices in the component indexed by Is and Ls, l the number in the component indexed by the Js and Ms, and $t \geq 1$ the number in both components,

$$P\{A \cap B\} = (1-s)^{k+l-t} \geq ((1-s)^l + s^l)(1-s)^k = P\{A\}P\{B\};$$

the inequality works because the extreme case is when $t = 1$ when the inequality asserts that $(1-s)^{l-1} \geq s^l + (1-s)^l \, \forall l$ which is equivalent to $(1-s)^{l-1} \geq s^{l-1}$ i.e that $s \leq \frac{1}{2}$.

We can now prove (@). As in Theorem 3.1 we first suppose C_1 and C_2 are edge-disjoint. Using the remarks in the previous paragraph, we have that

$$\mathbf{E}(I_{a_1} \ldots I_{a_i} L_{b_1} \ldots L_{b_l} J_{c_1} \ldots J_{c_j} M_{d_1} \ldots M_{d_m})$$

$$\geq \mathbf{E}(I_{a_1} \ldots I_{a_i} L_{b_1} \ldots L_{b_l}) \mathbf{E}(J_{c_1} \ldots J_{c_j} M_{d_1} \ldots M_{d_m}).$$

By the argument in Theorem 3.1 in turn, using the fact that there can be cycles on the left-hand side not on the right-hand side, and that in such a cycle an indicator can be forced to be same-same, or blue-blue, because all the other edges in the cycle are, we see that this is

$$\geq \mathbf{E}(I_{a_1} \ldots I_{a_i} L^*_{b_1} \ldots L^*_{b_l}) \mathbf{E}(J_{c_1} \ldots J_{c_j} M^*_{d_1} \ldots M^*_{d_m}).$$

Finally, it remains to pass from this back to the expression we want to have at the end of the proof, namely

$$\mathbf{E}\left(I_{a_1} \ldots I_{a_i} L^*_{b_1} \ldots L^*_{b_l} J_{c_1} \ldots J_{c_j} M^*_{d_1} \ldots M^*_{d_m}\right).$$

but of course on the surface the argument of the previous paragraph suggests that this may be somewhat greater than the last quantity considered. However, we will be able to get round this if we can show that

$$\mathbf{E}(I_{a_1} .. I_{a_i} L_{b_1} .. L_{b_l} J_{c_1} .. J_{c_j} M_{d_1} .. M_{d_m}) - \mathbf{E}(I_{a_1} .. I_{a_i} L_{b_1} .. L_{b_l}) \mathbf{E}(J_{c_1} .. J_{c_j} M_{d_1} .. M_{d_m})$$

$$\geq \mathbf{E}(I_{a_1} .. I_{a_i} L^*_{b_1} .. L^*_{b_l} J_{c_1} .. J_{c_j} M^*_{d_1} .. M^*_{d_m}) - \mathbf{E}(I_{a_1} .. I_{a_i} L^*_{b_1} .. L^*_{b_l}) \mathbf{E}(J_{c_1} .. J_{c_j} M^*_{d_1} .. M^*_{d_m}).$$

To prove this, recall we have just shown that $P\{A \cap B\} > P\{A\}P\{B\}$ when a component of the graph indexed by the Is and Ls intersects a component of the graph indexed by the Js and Ms, and note that this will happen more often when the Ls and Ms are present than when the L^*s and M^* are, as then there will be more intersections between C_1 and C_2.

The final step is to show that removing the edge-disjointness condition can only increase the difference. We used edge-disjointness to obtain the formula for $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$. However if they are not disjoint, the true value of $P\{C_1 \cap C_2\}$ will be at least as large as that suggested by the formula before, since that formula now double counts some of the edges and so includes a larger power of p, q or r than is in fact appropriate, so giving a value of $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$ lower than the correct one. So it suffices to show that (@) still holds. However the right-hand side of (@) is not affected by dropping the requirement that the subgraphs be edge disjoint, since C_1 and C_2^* are disjoint by definition and on the left-hand side there may still be additional correlation structure arising from cycles as before which will increase the expectation.

The case $q = 0$ is also handled by an argument similar to that used in Theorem 3.1. •

The proof may admit further generalisation. For example, it seems likely that if there are k colours, with same-same probabilities $p_{11} < p_{22} < ...p_{kk}$ and all same-different probabilities q, where $q < p_{jj} \forall j$, and where we look at the random variables S the number of same-same edges, T the number of edges between two vertices of the same colour from $s_2, ...s_k$, U the number of edges between two vertices of the same colour from $s_3, ...s_k$, and so on, that, provided a string of inequality conditions on the s_i hold, so as to ensure that when components of the graphs indexed by different types of indicators meet, the joint probability is at least as large as the product of their individual probabilities, we will get an analogous result. However we have not studied this in detail. Note that as soon as we let the same-different probabilities differ, the situation becomes more complex.

The argument of Theorem 3.10 depends on the assumption that $s \leq 1/2$. It is natural to investigate whether there is a simple example with $r > p$ and $s > 1/2$ for which the result fails, to see if the condition is genuinely necessary, as opposed to just being a limitation of our method. However such an example does not seem entirely straightforward to find. Of course the above proof will fail, since some colourings will give the wrong inequality

in (@) but it is not clear if these might be outweighed by the cases where the inequality goes the right way. The simplest possible example, by the results in Chapter 2 and earlier in this chapter is when we join a path of length one to a path of length two in the middle.

Lemma 3.11 *In ${}_sG_{p,q,r}$, the probability of this configuration is always at least as large as the product of the individual probabilities.*

Proof. This is a simple exercise in conditioning on the colours of the vertices; we have that the joint probability minus the individual probabilities is (by computer simplification and factorisation)

$$(1-s)\,s\,(r(1-s)+q+sp)\,(q-r+rs-2\,sq+sp)^2$$

which is clearly positive as required. •

Also, if we look at $C_1 = 1-2-3$ and $C_2 = 1-3$, then it seems on the basis of numerical trials that in any ${}_sG_{p,q,r}$ with $p, r > q$ we have $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$, though we do not have a complete proof here. Similarily a 4-cycle is not less likely than its individual edges by Theorem 2.27; and again numerical experiments suggest it is not less likely than the two paths of length 2 either. Lacking a clear method with which to attack this question in general, we leave it open.

3.7 Several colours, switches likelier than non-switches

We next consider what happens when there are several colours and the same-different probabilities are large but the same-same probabilities are small. The heuristic here is that with many colours there should be few same-same edges in both cycles and so it seems possible that we might yet get $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$. However we shall show that some care will be needed over the statement of any such result, using the following lemma.

Lemma 3.12 *Suppose we have a path of length $(n-1)$, $1-2-\ldots-n$ and have $k \geq 2$ colours. We assign to each of the vertices one of the k equiprobable colours, except that we insist that i and $i+1$ are of different colours at each stage, so that vertices $2,\ldots n$ are effectively being assigned one of $k-1$ colours equiprobably. Then the probability that the vertices 1 and n are the same colour is*

$$\frac{(k-1)^{n-2}-(-1)^{n-2}}{k(k-1)^{n-2}}.$$

Proof. Let a_n be the probability in question. Then by conditioning on the colour of vertex $(n-1)$ we have the obvious recurrence relation

$$a_n = a_{n-1}.0 + \frac{1-a_{n-1}}{k-1} \Leftrightarrow (k-1)a_n + a_{n-1} = 1,$$

with initial condition $a_3 = \frac{1}{k-1}$. It is then easy to check that the formula given is the solution of this recurrence. •

Theorem 3.13 *Suppose we have an RRC model with k equiprobable colours, with all same-same probabilities equal to p and all same-different probabilities equal to q, where $p < q$. Then there is a neighbourhood $q \in (\alpha, \alpha+\epsilon)$ in which $P\{C_1 \cap C_2\} \leq P\{C_1\}P\{C_2\}$ if the newgirth of the two subgraphs is odd, and $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ if the newgirth of the two subgraphs is even.*

Proof. Let S_1 be the number of edges of C_1 where the two end vertices are of **different** colours, and S_2 and S_2^* be similarily defined for C_2 and C_2^*. By arguments similar to those in Theorem 3.5 and 3.7, it suffices to prove

$$\mathbf{E}(I_{a_1} I_{a_2} .. I_{a_m} J_{b_1} ... J_{b_l}) \geq \mathbf{E}(I_{a_1} I_{a_2} ... I_{a_m} J^*_{b_1} ... J^*_{b_l})$$

for all choices of the I_i, J_j and J_j^*; here the I_i are indicators of whether the ith edge of C_1 is a switch, the J_j indicators of whether the jth edge of C_2 is a switch, and the J_j^* indicators of whether the jth edge of C_2^* is a switch. Again the expectation of a product of indicators is equal to the probability that the indicators are all 1, that is, that there is a switch of colour at each stage, and again correlation structure arises only if there are cycles amongst the indicators.

As in the proof of Theorem 3.5, we look at the first term for which the two sides differ, which is that term of the Taylor series where the power of θ is the newgirth of the two subgraphs, and then take θ small enough to obtain the statement of the theorem. We thus look at cycles whose length is the newgirth; then there is some cycle on the left-hand side which is not present on the right hand side. On the right-hand side, the probability all the indicators are 1 is $\left(1-\frac{1}{k}\right)^n$, since there is no cycle there and so at each stage the only question is whether the next vertex is a different colour from the present one. On the other hand, on the left-hand side, where there is a cycle, we must use Lemma 3.12 to close the cycle, and so we see that the probability is

$$\left(1-\frac{1}{k}\right)^{n-1}\left(1-\frac{(k-1)^{n-2}-(-1)^{n-2}}{k(k-1)^{n-2}}\right)$$

hence the left-hand side is greater than the right-hand side if and only if

$$\left(1 - \frac{(k-1)^{n-2} - (-1)^{n-2}}{k(k-1)^{n-2}}\right) > \left(1 - \frac{1}{k}\right)$$

$$\Leftrightarrow 1 > \frac{(k-1)^{n-2} - (-1)^{n-2}}{(k-1)^{n-2}} \Leftrightarrow n \text{ is even}$$

giving the result. •

Again it may be that the assumption of equiprobable colours can be relaxed at the expense of introducing more complex formulae. Of course the above result only describes behaviour in a neighbourhood of $p = q$; it may be true that, for q sufficiently greater than p in some sense and sufficiently many colours, the inequality $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ holds; however it is not clear what such a statement should be.

3.8 The role of the FKG inequalities.

We now turn as promised earlier to the question of how Theorem 3.1 ties up with the FKG inequalities and related matters. We first recall a definition which will allow us to state the FKG inequality;

Definition 3.3 *A function f from the subsets of a finite set to the real numbers is* **non-decreasing** *if $A \subset B \Rightarrow f(A) \leq f(B)$, and an event is non-decreasing if its indicator function is non-decreasing.*

Thus, for example, the event that all of some specified collection of edges arises in a RRC graph is clearly an increasing event.

Theorem 3.14 *Suppose that μ is a probability measure on the subsets of some finite set S, satisfying the inequality*

$$\mu\{x \cup y\}\mu\{x \cap y\} \geq \mu\{x\}\mu\{y\} \, \forall x, y \in L$$

(such a μ is said to be **log-supermodular***). Then if f and g are non-negative non-decreasing functions on ve*

$$\sum_{x \in L} \mu(x) f(x) \sum_{x \in L} \mu(x) g(x) \leq \sum_{x \in L} \mu(x) f(x) g(x).$$

That is, $\mathbf{E}(f(x))\mathbf{E}(g(x)) \leq \mathbf{E}(f(x)g(x))$.

Proof. This is standard; see e.g [B2], Theorem 19.5. •

Log-supermodularity is often satisfied, as the next example shows;

Lemma 3.15 *Suppose* $\mathbf{s} = (s_1, s_2, ...s_k)$ *with* $\sum_{i=1}^{k} s_i = 1$ *and* $s_i \geq 0 \forall i$. *Define, for* $A \subset \{1, 2, ...k\}$ $\mu(A) = \prod_{i \in A} s_i \prod_{i \notin A}(1 - s_i)$ *so that elements of* $\{1, 2, ...k\}$ *are in* A *independently, with* $P\{i \in A\} = s_i$. *Then* μ *is log-supermodular; more precisely,* $\mu\{A \cup B\}\mu\{A \cap B\} = \mu\{A\}\mu\{B\}$.

Proof. This is again standard, and easy to see by considering the contribution of each element i of $\{1, 2...k\}$ to both sides. •

We first explain how the FKG inequality is used classically to obtain the analogue of Theorem 3.1.

Theorem 3.16 *Suppose* μ *is a measure for which the FKG inequality holds (e.g* μ *is log-supermodular). Then, if* A *and* B *are events whose indicator functions are non-decreasing, we have*

$$P\{A \cap B\} \geq P\{A\}P\{B\}.$$

and if C *is an nondecreasing event and* D *is a nonincreasing event (that is, the indicator function of* D *is nonincreasing) we have*

$$P\{C \cap D\} \leq P\{C\}P\{D\}.$$

Proof. This again is standard; the first claim is proven by applying the FKG inequality when f is the indicator function of A and g is the indicator function of whether the set contains all the elements of B (so that f and g are clearly non-negative and increasing), and noting that the expectation of an indicator variable is the probability that it is 1.

For the second part, let f be the indicator function of C and g the indicator function of D, so that f is nondecreasing and g nonincreasing; then f and $h = \sup_{x \in L}(g) - g$ are both nondecreasing and positive. Applying the FKG inequality to them and tidying up we get that $P\{C \cap D\} \leq P\{C\}P\{D\}$ as required. •

The result of Harris which was mentioned in the remarks after Theorem 3.1 is an immediate corollary of Theorem 3.16 since the event that all the edges in some set arise is clearly an increasing event.

We now show that the FKG inequality fails for μ the probability measure in $G_{p,q}$ with $p > q$. Thus, although $P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}$ still, by

Theorem 3.1, the classical method of proof cannot be used. More precisely, we shall exhibit three subgraphs C_1, C_2 and C_3 for which, although $p > q$,

$$P\{C_3 \cap (C_1 \cup C_2)\} \leq P\{C_3\}P\{C_1 \cup C_2\}.$$

(Note in passing that if it were not for the fact that we only consider events of the form "all edges in a certain set arise" in Theorem 3.1, the previous inequality would hold more generally. Indeed we would have, using repeatedly the easy fact that $C_1 \cup C_2 = (C_1^c \cap C_2^c)^c$,

$$P\{C_3 \cap (C_1 \cup C_2)\} = P\{C_3\} - P\{C_3 \cap C_1^c \cap C_2^c\}$$

$$\leq P\{C_3\} - P\{C_3\}P\{C_1^c \cap C_2^c\} \text{ by Theorem 3.1}$$

$$= P\{C_3\}P\{C_1 \cap C_2\}).$$

Let C_1 be the graph on vertices $1, 2, 3, 4$ with edges $1-2$ and $2-4$, C_2 be the graph on these vertices with edges $1-3$ and $3-4$, and C_3 be the edge $2-3$ (note these three subgraphs are edge-disjoint). We note a lemma which will be used again in section 3.11.

Lemma 3.17 *Let A be the triangle $1-2-3-1$ and B be the triangle $1-2-4-1$. Then in $G_{p,q}$, with $\alpha = (p+q)/2$ and $\beta = (p-q)/2$ as before,*

$$P\{A \cap B\} = \frac{p}{8}(p^4 + 2p^2q^2 + 4pq^3 + q^4)$$

$$= \frac{(\alpha+\beta)(\alpha^2+\beta^2)^2 + (\alpha-\beta)(\alpha^2-\beta^2)^2}{2}.$$

Proof. Considering whether 1 and 2 are the same or different colours, and using Theorem 2.6, we have

$$\frac{p}{2}\left(\left(\frac{p+q}{2}\right)^2 + \left(\frac{p-q}{2}\right)^2\right)^2 + \frac{q}{2}\left(\left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2\right)^2$$

and this is easily checked to be equal to both expressions. •

Then in $G_{p,q}$ using Theorems 2.4 and 2.6 it is easy to see that

$$P\{C_3 \cap \{C_1 \cup C_2\}\} = P\{C_3 \cap C_1\} + P\{C_3 \cap C_2\} - P\{C_1 \cap C_2 \cap C_3\}$$

$$= 2\left(\frac{p+q}{2}\right)^3 - \frac{p}{8}\left(p^4 + 2p^2q^2 + 4pq^3 + q^4\right).$$

However

$$P\{C_3\}P\{C_1 \cup C_2\} = P\{C_3\}\left(P\{C_1\} + P\{C_2\} - P\{C_1 \cap C_2\}\right)$$

$$= \frac{p+q}{2}\left(2\left(\frac{p+q}{2}\right)^2 - \left(\frac{p+q}{2}\right)^4 - \left(\frac{p-q}{2}\right)^4\right)$$

whence it is easy to show that

$$P\{C_3 \cap (C_1 \cup C_2)\} - P\{C_3\}P\{C_1 \cup C_2\} = -\frac{(p+q)^2(p-q)^3}{16} < 0 \text{ for } p > q.$$

In particular, the measure is not log-supermodular. Similar more general examples along the same lines could be constructed by considering a rhombus with all sides paths of length k, letting C_1 consist of two adjacent sides of the rhombus and C_2 consist of the other two sides, so that $C_1 \cap C_2$ is the rhombus, and letting C_3 be a path of length l between two opposite vertices, one of them in C_1 and the other in C_2; in this case similar reasoning shows that

$$P\{C_3 \cup (C_1 \cap C_2)\} - P\{C_3\}P\{C_1 \cup C_2\} = -\frac{(p+q)^{2k}(p-q)^{2k+l}}{2}$$

which again is negative for $p > q$. So the phenomenon is not purely one that is restricted to small subgraphs. It seems likely that it has more to do with the fact, that if C_1 and C_2 are present, this implies that there is a high probability that the two end vertices of C_3 are the same colour and so $P\{C_1 \cap C_2 \cap C_3\}$ is larger than one might otherwise expect, since the presence of C_3 is positively correlated with the presence of C_1 and C_2.

Note that the failure of the FKG inequalities must reflect the lack of independence in our models, as Lemma 3.15 makes it clear that, for these models, inhomogeneity alone is not an obstruction to the results. That the FKG inequality can fail when independence no longer holds has been noted before; for example, in [ASE] it is noted that if A is a fixed subset of $\{1, 2, ...k\}$ and A_i, $1 \leq i \leq m$ are random subsets of $\{1, 2, ...k\}$ where $P\{l \in A_i\} = p_l$, then

$$P\{A \cap A_i \neq \emptyset \forall i\} \geq \prod_{1 \leq i \leq m} P\{A \cap A_i \neq \emptyset\}$$

but that this can fail if instead the A_i are random r-subsets of $\{1, 2, ...k\}$.

3.9 Some remarks on the Janson inequality

A consequence of the FKG inequality which is commonly used in the classical theory is the so-called **Janson inequality**;

Theorem 3.18 *Suppose* $\{A_i\}$ *for* $i \in \{1, 2, ...n\}$ *are events in a probability space such that, for* $S \subset \{1, 2, ...n\}$, *we have*

$$1. \forall i \, and \, S \, with \, i \notin S \; P\{A_i \mid \cap_{j \in S} A_j^c\} \leq P\{A_i\}.$$

$$2. \forall i \neq j \, and \, S \, with \, i, j \notin S \; P\{A_i \cap A_j \mid \cap_{k \in S} A_k^c\} \leq P\{A_i \cap A_j\}.$$

Let $M = \prod_{i=1}^n P\{A_i^c\}$; *then, if* $P\{A_i\} \leq \epsilon \forall i$ *and* $\Delta = \sum P\{A_i \cap A_j\}$ *where the sum is taken over pairs of events which are dependent but not identical, we have*

$$M \leq P\{\cap_{i=1}^n A_i^c\} \leq M e^{\frac{\Delta}{2(1-\epsilon)}}$$

In particular, if $\epsilon = o(1)$ *and* $\Delta = o(1)$ *we get the asymptotic formula*

$$P\{\cap A_i^c\} \sim M.$$

Proof. [ASE] gives a proof due to Boppana and Spencer, which is somewhat more elementary than Janson's original one. •

In our applications, of course, the events A_i will be that some set of edges arises. In particular, they will be increasing events, so both conditions in the Janson inequality would follow if the FKG inequalities held, since if A_i is an increasing event A_i^c is a decreasing event. However our counterexample to the FKG inequalities above implies that even the first condition in the Janson inequalities does not hold in our models. Indeed recall that we exhibited three subgraphs such that

$$P\{C_3 \cap (C_1 \cup C_2)\} \leq P\{C_3\} P\{C_1 \cup C_2\}.$$

Now, by elementary probability theory, we have

$$P\{C_3 \cap (C_1 \cup C_2)\} \leq P\{C_3\} P\{C_1 \cup C_2\}$$

$$\Leftrightarrow P\{C_3 \cap (C_1 \cup C_2)^c\} \geq P\{C_3\} P\{(C_1 \cup C_2)^c\}$$

$$\Leftrightarrow P\{C_3 \cap C_1^c \cap C_2^c\} \geq P\{C_3\} P\{C_1^c \cap C_2^c\}$$

$$\Leftrightarrow P\{C_3 \mid C_1^c \cap C_2^c\} \geq P\{C_3\}.$$

and so for the C_i as in the previous section, the first condition in the Janson inequalities fails to hold for **any** $p > q$.

The failure of the Janson inequalities in our setup is is a pity since in the classical model this theorem is an important tool for estimating joint probabilities of events. For example, since we lack a tool to prove that the probability of all of some collection of events is close to the product of their probabilities, if we have some random variable which is a sum of identically distributed indicators, we will have difficulty proving that the r-th moments of the sum are close to the r-th powers of the mean we want to converge to; this suggests that the Poisson paradigm (which means essentially the idea that the probability of rare events can be approximated, for large n by a Poisson distribution) may be less widely applicable in our models than classically.

Of course in practice the Janson inequalities are often used as an asymptotic tool, and the above argument does not rule out the possibility that it may still be possible to get good approximations to the probability of an intersection of events by the product of their probabilities; and we shall see some partial evidence to support this idea in the section 3.11. However a more sophisticated argument than classically will be needed, and it is not clear how generally such ideas can be got to work.

3.10 Are S_1 and S_2 associated random variables?

We begin by recalling the following definition.

Definition 3.4 *Random variables S and T are* **associated** *if*

$$\mathbf{E}(f(S)g(T)) \geq \mathbf{E}(f(S))\mathbf{E}(g(T))$$

for all bounded increasing functions f and g.

This inequality is (rather confusingly) sometimes called the FKG inequality; we will call it FKG(2) to distinguish it from what we called the FKG inequality. Recall that in Theorem 3.3 above we showed that

$$\mathbf{E}(S_1^r S_2^s) \geq \mathbf{E}(S_1^r)\mathbf{E}(S_2^s).$$

This suggests the possibility at least that the random variables S_1 and S_2 are associated, but this too turns out to be false. To get a counterexample,

note that Theorem 3.3 implies the special case of FKG(2) where f and g are both of the form $x \mapsto x^r$; consequently, it holds for any pair of increasing functions f and g which can both be expressed as a limit of polynomials with non-negative coefficients. However this does not cover all increasing functions; for example, any such function, being a limit of a series whose terms are convex functions is itself a convex function. Thus we try non-convex functions to get a counterexample and in fact some simple ones work. For example, if C_1 is the cycle $1-2-3-4-5-1$ and $C_2 = 1-3-5-2-4-1$, and $f(x) = g(x) = 1 - e^{-x}$ which is clearly bounded and increasing but is concave we have

$$\mathbf{E}(f(S_1)f(S_2)) \geq \mathbf{E}(f(S_1))\mathbf{E}(f(S_2)) \Leftrightarrow \mathbf{E}\left(e^{-(S_1+S_2)}\right) \geq \mathbf{E}\left(e^{-S_1}\right)\mathbf{E}\left(e^{-S_2}\right)$$

on simplifying. Considering the following cases;

1. All five vertices the same colour (2 of the 32 cases): $S_1 = 5$, $S_2 = 5$.

2. All vertices except one are the same colour (10 cases): $S_1 = 3$, $S_2 = 3$.

3. Three vertices are one colour, two vertices which are adjacent on C_1 are the other colour (10 cases): then $S_1 = 3$, $S_2 = 1$.

4. Three vertices are one colour, but the other two vertices, which are not adjacent on C_1 are the same colour (10 cases); S_1=1, S_2=3.

we see that

$$\mathbf{E}\left(e^{-(S_1+S_2)}\right) = \sum xP\{e^{-(S_1+S_2)} = x\} = \frac{e^{-10}}{16} + \frac{5e^{-6}}{16} + \frac{5e^{-4}}{8} \simeq 0.012224$$

whereas

$$\mathbf{E}(e^{-S_1})\mathbf{E}(e^{-S_2}) = (\sum xP\{e^{-S_1} = x\})^2 = (\frac{e^{-5}}{16} + \frac{5e^{-3}}{8} + \frac{5e^{-1}}{16})^2 \simeq 0.0214$$

as required. In fact we can do a similar calculation for the same two cycles with the functions f and g both being $x \to x^{\frac{1}{2}}$ and we again find that the random variables S_1 and S_2 are not associated.

The above discussion leaves open the possibility that FKG(2) might still hold if we only look at bounded increasing **convex** functions f and

g. However it is fairly clear that not all increasing convex functions arise as limits of polynomials with positive coefficients (for example, those with bad differentiability properties), so something more would have to be said. The notion of association arising when we restrict the increasing functions f and g to be convex does not seem to have been studied in the literature on association, and it is far from obvious if such a property would be useful.

3.11 Some exact results on numbers of 3-cycles

It is of some interest to do some exact calculations for N_3 the number of 3-cycles. This section is partly motivated by the failure of the Janson inequalities discussed above, since in the classical model they are a standard tool for estimating the probability that a graph is triangle-free (see [S]). Some of our results will be seen to offer some support for the notion that, although the Janson inequalities fail in simplistic form, something similar seems to work at least some of the time. We start by giving exact formulae for the expectation and variance of the number of 3-cycles.

Theorem 3.19 *Let N_3 be the number of 3-cycles in $G_{p,q}$. Then*

$$\mathbf{E}N_3 = \binom{n}{3}\left(\left(\frac{p+q}{2}\right)^3 + \left(\frac{p-q}{2}\right)^3\right)$$

and $Var(N_3)$, the variance of N_3 is equal to

$$-\frac{pn(n-1)(n-2)(18np^3q^2 - 24npq^3 - 6nq^4 - 6np^4 - 12np^2q^2)}{96}$$

$$-\frac{pn(n-1)(n-2)(27npq^4 + 3np^5 - 48p^3q^2 - 4p^2 - 12q^2)}{96}$$

$$-\frac{pn(n-1)(n-2)(72pq^3 + 36p^2q^2 - 8p^5 + 18p^4 - 72pq^4 + 18q^4)}{96}$$

For sufficiently large n, $Var(N_3)$ is greater than or less than its classical value according as $p > q$ or $< q$.

Proof. Since N_3 is the sum of indicator variables of whether or not each 3-cycle is present, using the formula for the probability of a cycle in Theorem 2.6 and the linearity of expectation, we have

$$\mathbf{E}(N_3) = \binom{n}{3}\left(\left(\frac{p+q}{2}\right)^3 + \left(\frac{p-q}{2}\right)^3\right)$$

For the variance, we need to obtain $\mathbf{E}(N_3^2)$. Now N_3 is a sum of indicator variables of whether the potential cycles arise. If two 3-sets have no vertices in common, as $\binom{n}{3}\binom{n-3}{3}$ of the $\binom{n}{3}^2$ pairs of 3-sets do, or one element in common, as $3\binom{n}{3}\binom{n-3}{2}$ do, the existence of the two cycles is independent by Lemma 2.1 and the remarks before Theorem 3.1. If they are the same, as $\binom{n}{3}$ pairs are, the joint probability is the probability of one of them. The only other case is if they have one common edge when by Lemma 3.17 the joint probability is $p(p^4 + 2p^2q^2 + 4pq^3 + q^4)/8$. Thus

$$E\left(N_3^2\right) = \binom{n}{3}\left(\binom{n-3}{3} + 3\binom{n-3}{2}\right)\left(\left(\frac{p+q}{2}\right)^3 + \left(\frac{p-q}{2}\right)^3\right)^2$$

$$+3\binom{n}{3}\binom{n-3}{1}\frac{p\left(p^4 + 2p^2q^2 + 4pq^3 + q^4\right)}{8} + \binom{n}{3}\left(\left(\frac{p+q}{2}\right)^3 + \left(\frac{p-q}{2}\right)^3\right).$$

As $\binom{n-3}{3} + 3\binom{n-3}{2} = \binom{n}{3} - 3\binom{n-3}{1} - 1 = \binom{n}{3} - (3n-8)$, we have that (temporarily using α and β to simplify writing the expressions)

$$\mathrm{Var}_{p,q}(N_3) = \mathbf{E}(N_3^2) - \mathbf{E}\left(N_3\right)^2$$

$$= \binom{n}{3}^2(\alpha^3+\beta^3)^2 - (3n-8)\binom{n}{3}(\alpha^3+\beta^3)^2 - \binom{n}{3}^2(\alpha^3+\beta^3)^2$$

$$+3\binom{n}{3}\binom{n-3}{1}\frac{(\alpha+\beta)(\alpha^2+\beta^2)^2 + (\alpha-\beta)(\alpha^2-\beta^2)^2}{2} + \binom{n}{3}(\alpha^3+\beta^3)$$

$$= -(3n-8)\binom{n}{3}(\alpha^3+\beta^3)^2$$

$$+3\binom{n}{3}\binom{n-3}{1}\frac{(\alpha+\beta)(\alpha^2+\beta^2)^2 + (\alpha-\beta)(\alpha^2-\beta^2)^2}{2} + \binom{n}{3}(\alpha^3+\beta^3)$$

$$= -\frac{pn(n-1)(n-2)(18np^3q^2 - 24npq^3 - 6nq^4 - 6np^4 - 12np^2q^2)}{96}$$

$$-\frac{pn(n-1)(n-2)(27npq^4 + 3np^5 - 48p^3q^2 - 4p^2 - 12q^2)}{96}$$

$$-\frac{pn(n-1)(n-2)(72pq^3 + 36p^2q^2 - 8p^5 + 18p^4 - 72pq^4 + 18q^4)}{96}$$

by computer simplification. Thus

$$\frac{384(\mathrm{Var}_{p,q}(N_3) - \mathrm{Var}_{\alpha}(N_3))}{n(n-1)(n-2)(q-p)^3} = \left(9p^3 + 3q^3 + 27pq^2 + 9p^2q - 18p^2 - 6q^2 - 24pq\right) n$$

$$+ \left(-8q^3 - 72pq^2 + 18q^2 - 24p^2q + 72pq - 8 + 54p^2 - 24p^3\right) (*).$$

For the last claim, note that the coefficient of n on the right-hand side is always non-positive as

$$9p^3 + 3q^3 + 27pq^2 + 9p^2q - 18p^2 - 6q^2 - 24pq$$

$$= 9\left(p^3 - p^2\right) + 9\left(p^2q - p^2\right) + 3\left(q^3 - q^2\right) + 24\left(pq^2 - pq\right) + 3(pq^2 - q^2)$$

and each individual bracket is clearly non-positive (they can all be zero if $p = q = 1$ or $p = q = 0$, but then of course the constant term is also 0). Thus for n sufficiently large, the variance with $p > q$ is always greater than or equal to its classical value, and for $p < q$ it is always less than or equal to its classical value. •

Note that whilst Theorem 3.4 says that $\mathbf{E}_{p,q}\left(N_3^i\right) \geq \mathbf{E}_{\alpha}\left(N_3^i\right)$ for $p > q$, this of course does not imply $\mathrm{Var}_{p,q}N_3 \geq \mathrm{Var}_{\alpha}N_3$ for $p > q$. Indeed if $p = 0.999$ and $q = 0.998$ the right-hand side of $(*)$ is $-0.08376616n + 8.1790642$ which is only negative for $n \geq 98$. However, this does show that with $p > q$ the variance is greater than classically for large enough n.

Theorem 3.19 has implications for the question of when a random graph is **triangle free** that is, there are no triangles (3-cycles) present.

Theorem 3.20 *In $G_{p,q}$, let $p \sim c/n$ and $q \sim d/n$ where c and d are constant. Then*

$$\lim_{n\to\infty} P\{G \textit{ is triangle free}\} = e^{-\frac{(c^3+3cd^2)}{24}}.$$

Proof. We first notice that the relevant probability is the probability of the two events A that there is no monochrome triangle and B that there is no polychrome triangle. Next note the probability that the event happens equals the probability that it happens conditional on the number of reds and blues both being $n/2 + o(n^{1/2+\epsilon})$, since the latter happens with probability tending to one. Given the numbers of reds and blues, of course, the events A and B are independent, and do not depend (in the limit) on the variability in the numbers of reds and blues.

We estimate A first. Recall that the classical calculation of the probability that $G_{\frac{a}{n}}$ is triangle-free is fairly robust to the exact form of the probability (see [S] for discussion of this); in particular, as it depends only on the fact that $\alpha \sim \frac{a}{n}$, it is not sensitive to the slight variability in the number of reds and blues and so

$$\lim_{n\to\infty} P\{\text{no monochromatic triangle}\} = \left(\lim_{n\to\infty} P\{N_3\left(G\left(\frac{n}{2}, \frac{2a}{n}\right)\right) = 0\}\right)^2$$

$$= \left(e^{-\frac{a^3}{6}}\right)^2 \text{ by the classical calculation} = e^{-\frac{c^3}{24}} \text{ as } c = 2a.$$

Thus it suffices to show that $P\{\text{no polychromatic triangle}\} \sim e^{-cd^2/8}$.

We use the Janson inequalities. Each potential polychromatic triangle arises with probability pq^2. The two assumptions required for the Janson inequalities hold for the potential polychromatic triangles, since we have conditioned on the number of reds and blues, so we need only show that (in the terminology of Theorem 3.18) Δ and ϵ are $o(1)$ and that $M \sim e^{-cd^2/8}$.

To do this, note first that as p and q are both $\sim c\kappa/n$ for some constant κ, we have $\epsilon = o(1)$ as required (indeed it is $O(n^{-3})$). To show that Δ is $o(1)$, we need only consider the case when the two polychrome triangles have one edge in common, as otherwise they are independent or identical. In this case, there are about $(n/2)^4$ choices of the vertices in the two triangles, and since the probability of the two triangles arising is some constant divided by n^5 we get that $\Delta = o(1)$ as required.

Finally, to apply Theorem 3.18 we note that, since if the first vertex is from the about $n/2$ reds the other two vertices are chosen from the about $n/2$ blues, and similarily if the first vertex is blue; again the slight variability in the numbers of reds and blues is not important, and thus

$$M = (1 - \frac{cd^2}{n^3})^{2.n/2.\binom{n/2}{2}} \sim e^{-cd^2/8}. \bullet$$

It may be possible to use similar techniques to obtain insight into the probability that the graph contains no subgraph isomorphic to some fixed graph H for at least certain other types of H, and possibly other questions which are classically solved with the Janson inequalities. However, we have not examined this in detail.

4 Manifestation of correlation structure in the number of edges

4.1 The number of edges; introductory remarks

In the previous two chapters we concentrated primarily on questions about the probability of particular patterns of edges, such as trees or cycles. In this chapter and the next we instead consider the simpler random variable $\mathcal{E}$, the number of edges in the graph (we shall sometimes write $\mathcal{E}_\alpha$ or $\mathcal{E}_{p,q}$ if we wish to emphasise the model in which we are working). We also consider some closely related random variables, such as the degrees of the vertices. Often we shall compare moments $\mathbf{E}(\mathcal{E}^r)$ in the new and classical models, or look at several moments simultaneously via a generating function. In Chapter 5 we shall continue this general line of enquiry by looking in some detail at large deviation principles for our models (which depend on all moments through the moment generating function).

In the classical model $\mathcal{E}_\alpha \sim \text{Bin}(n(n-1)/2, \alpha)$, and it is consequently as well (or ill!) understood as such random variables are. However in our models the distribution will usually be more complex. For example, even in $G_{p,q}$ the probability function is the following rather messy expression which does not seem to have any simplification. We adopt the convention that any binomial coefficient $\binom{a}{b}$ is zero unless $a \geq b$ are non-negative integers).

Lemma 4.1 *$P\{\mathcal{E}_{p,q} = k\}$ is given by*

$$\sum_{i=0}^{n}\sum_{j=0}^{k}\binom{\frac{n^2-n}{2}-i(n-i)}{j}p^j(1-p)^{\frac{n^2-n}{2}-i(n-i)-j}\binom{i(n-i)}{k-j}q^{k-j}(1-q)^{i(n-i)-k+j}\frac{\binom{n}{i}}{2^n}.$$

Proof. Conditional on their being i reds, which happens with probability $\binom{n}{i}/2^n$, there are $i(n-i)$ same-different edges and hence $n(n-1)/2 - i(n-i)$ same-same edges. Thus to get a total of k successes, we must have, for some $0 \leq j \leq k$, $k-j$ successes in the $i(n-i)$ independent trials with success probability q and j successes in the $n(n-1)/2 - i(n-i)$ independent trials with probability q; the result follows from the form of the probability function of the binomial distribution. •

The analogous formulae in more complex RRC models would be even worse. Thus we concentrate on other aspects about which rather more can be said.

4.2 Generalisations of independence for the edges.

Of course $\mathcal{E} = \sum_{1 \le i < j \le n} X_{ij}$ where $X_{ij} = 1$ if edge $i - j$ is present and 0 otherwise (we put $X_{ii} = 0 \; \forall i$). Classically the X_{ij} are independent, but Theorem 2.17 shows that this will be true here only for trivial models. We thus ask what properties generalising independence the X_{ij} have.

We first note that the X_{ij} are not **exchangeable**. Recall discrete random variables $X_1, ..X_n$ are exchangeable if and only if for all n and permutations π of $\{1, 2...n\}$ we have

$$P\{X_1 = x_1, ...X_n = x_n\} = P\{X_{\pi(1)} = x_1, X_{\pi(2)} = x_2, ...X_{\pi(n)} = x_n\}$$

and that an infinite such sequence is exchangeable if and only if all its finite subsequences are. For example, i.i.d random variables are exchangeable. See the survey [A] for more information about exchangeability.

Lemma 4.2 *Suppose the indicators X_{ij} of the edges in an RRC model are exchangeable. Then there are at most three vertices or the model is trivial in the sense of Theorem 2.17.*

Proof. If there are at least four vertices and the model is not TID, then letting $1, 2, 3, 4$ be some of the vertices, the events A, that the edges $1-2$ and $3-4$ both arise, and B, that the edges $1-2$ and $1-3$ arise, have different probabilities, contradicting exchangeability. Thus we now assume the model is TID. If there are more than four vertices, we consider the events A that the cycle $1-2-3-4-1$ arises, and B that the tree $1-2-3-4-5$ arises. By exchangeability and TID, these both have probability α^4; but by Theorem 2.27 if a cycle of even length has the same probability as classically, the model is trivial as required.

On the other hand, it is easy to check that for three vertices or less, the X_{ij} are exchangeable in all RRC models. Thus it only remains to show that if the indicators are exchangeable when there are four vertices and the model is TID, then the model is trivial. For this, we note that exchangeability implies that the probability of a 4-cycle is equal to the probability of a 3-cycle C with an extra edge attached to one vertex v say; this probability is

$$\sum_{1 \le i \le k} s_i P\{C \mid c(v) = i\} \sum_{j=1}^{k} p_{ij} s_j = \alpha P\{C\}$$

since the $\sum_{j=1}^{k} p_{ij} s_j$ are equal when $s_i \neq 0$ by the TID assumption; it also implies that $P\{C\} = \alpha^3$, since the latter is the probability of the path

$1-2-3-4$ by exchangeability and the fact that the model is TID. Hence together these assumptions force the probability of a 4-cycle to be the same as classically, and again the result follows from Theorem 2.27. •

Lemma 4.2 is unsurprising since De Finetti's theorem, a key result in exchangeability theory says that an infinite sequence is exchangeable if and only if it is a mixture of i.i.d sequences ([A], section 2); that is, conditional on some (possibly vector-valued) parameter, the variables are i.i.d. However here, whilst independence arises conditional on the colours of all the vertices, they are not then identically distributed unless the model is trivial. The X_{ij} do however have the property that for all $1 \leq i, j \leq n$

$$P\{X_{\pi(i)\pi(j)} = x_{ij} \forall i, j\} = P\{X_{ij} = x_{ij} \forall i, j\};$$

informally, permuting the labels of the vertices leaves the distribution of the degrees unchanged. This property is termed **weak exchangeability** in [A]; the general structure theorem 14.21 proven there, namely that any weakly exchangeable array is of the form $X_{ij} = g(\psi, \xi_i, \xi_j, \lambda_{i,j})$, where g is any function such that $g(a, ., ., d)$ is symmetric for any (a, d) and ψ, ξ_i and $\lambda_{i,j}$ are uniform on $(0, 1)$, says nothing in this particular case which is not already obvious, but the property will be useful when we consider a central limit theorem for $\mathcal{E}$, when this property will be seen to be equivalent to a property we will require then.

Another direction in which one can generalise the notion of independence is to dissociated random variables [BHJ];

Definition 4.1 *For a collection* Γ *of* k*-subsets* $i = \{i_1, ...i_k\}$ *of* $\{1, 2, ...n\}$*, random variables* I_i *for* $i \in \Gamma$ *are* **dissociated** *if the subsets of random variables*

$$(I_i,\, i \in A) \text{ and } (I_i,\, i \in B)$$

are independent whenever

$$\left(\bigcup_{i \in A} i\right) \cap \left(\bigcup_{i \in B} i\right) = \emptyset.$$

For $k = 1$ dissociation is the same as independence of the I_i, but for $k \geq 2$ it is (much) more general. Of course for us $k = 2$, the set Γ consists of all $n(n-1)/2$ potential edges of the graph, and the indicators are dissociated by Lemma 2.1. In fact a stronger property will sometimes hold;

Definition 4.2 *A dissociated family* $(I_j, j \in \Gamma)$ *of random variables, indexed by the edges of a graph* Γ *is* **strongly dissociated** *if also,* $\forall H \subset \Gamma$

$$| j \cap \cup_{i \in H} i | = 1 \Rightarrow I_j \text{ is independent of } (I_i : i \in H).$$

Lemma 4.3 *An RRC model is TID if and only if the indicators of the edges are strongly dissociated.*

Proof. If the model is TID, consider $P\{I_j = 1 \text{ and } I_i = 1 \forall i \in K \subset H\}$. If the $i \in K$ and j have no common vertex, the events are independent by Theorem 2.1; otherwise there is one vertex, v say, in common and

$$P\{I_j = 1 \text{ and } I_i = 1 \forall i \in K \subset H\}$$

$$= \sum_{l=1}^{k} P\{I_j = 1 \text{ and } I_i = 1 \forall i \in K \mid c(v) = l\} s_l.$$

$$= \sum_{l=1}^{k} P\{I_j = 1 \mid c(v) = l\} P\{I_i = 1 \forall i \in K \mid c(v) = l\} s_l.$$

But as the model is TID, $P\{I_i = 1 \mid c(v) = l\} = \alpha$ if $P\{c(v) = l\} \neq 0$ by Theorem 2.2, and so this is $P\{I_j = 1\} P\{I_i = 1 \forall i \in K\}$ as required.

Conversely, if the indicators are strongly dissociated, the probability of the path $1 - 2 - 3$ is (treating 2 as the common vertex between j the edge $1 - 2$ and i the edge $2 - 3$) the product of their probabilities which is its classical value; and as in Theorem 2.2, this implies the model is TID. •

Note that in general there are dissociated families of pairwise independent random variables which are not strongly dissociated. Dissociated and strongly dissociated variables are widely used in Poisson approximation theory ([BHJ]) and will be useful in our discussion of that topic for $\mathcal{E}$ later.

4.3 Expectation and variance of the number of edges.

We first consider the expected number of edges of $\mathcal{E}$, which is very easy to relate to its classical value.

Theorem 4.4 $\mathbf{E}\mathcal{E}_\alpha = \mathbf{E}\mathcal{E}$ *in an RRC model* $\Gamma(n, k, P, \mathbf{s})$.

Proof. Immediate from $\mathcal{E} = \sum_{1 \le i < j \le n} X_{ij}$, the fact $\mathbf{E}X_{ij} = \alpha$ and linearity of expectation. •

We next consider the variance $\mathrm{Var}(\mathcal{E})$. Since the variance of a sum of n independent Bernoulli random variables is maximised, as their probabilities p_i vary with $\sum_{i=1}^{n} p_i = n\alpha$ fixed, when all the p_i are equal, we might naively guess that $\mathrm{Var}(\mathcal{E})$ will be smaller in RRC models than classically. However it transpires that the correlation structure at least cancels out this effect.

Lemma 4.5 *Let X_i $(1 \le i \le n)$ be identically distributed and pairwise independent. Then $Var(\sum_{i=1}^{n} X_i) = n\,Var(X_i)$.*

Proof. Expanding, and using that the X_i are identically distributed,

$$\mathrm{Var}(\sum_{1 \le i \le n} X_i) = \mathrm{Cov}(\sum_{1 \le i \le n} X_i, \sum_{1 \le i \le n} X_i)$$

$$= n\mathrm{Cov}(X_i, X_i) + n(n-1)\mathrm{Cov}(X_i, X_j) = n\mathrm{Var}(X_i) + n(n-1)\mathrm{Cov}(X_i, X_j);$$

where $j \neq i$; but $\mathrm{Cov}(X_i, X_j) = 0$ as the X_i are pairwise independent so this is $n\mathrm{Var}(X_i)$ as required. •

Theorem 4.6 *In any TID RRC model $\Gamma(n, k, P, \mathbf{s})$, $Var(\mathcal{E})$ depends only on $\alpha = \sum s_i s_j p_{ij}$.*

Proof. $\mathcal{E} = \sum_{1 \le i < j \le n} X_{ij}$. Since the model is TID the X_{ij} are pairwise independent with the same distribution. The result follows by Lemma 4.5 taking the Y_i to be the indicators in the corresponding classical model. •

A similar argument bounds $\mathrm{Var}(\mathcal{E})$ below for any RRC model;

Theorem 4.7 *In any RRC model Γ, $Var(\mathcal{E}_\Gamma) \ge Var(\mathcal{E}_\alpha)$ with equality if and only if the model is TID.*

Proof. As in Theorem 4.5 $\mathrm{Var}\mathcal{E} = \sum_{1 \le i,j,k,l \le n} \mathrm{Cov}(X_{ij}, X_{kl})$. The covariances where X_{ij} and X_{kl} have no vertex in common, or are the same edge, are the same as classically, so we need only consider the cases where there is one common vertex. Then the two edges form a tree so by Theorem 2.22 the joint probability is no less than classically, with equality if and only if the model is TID. The result follows. •

One can of course prove Theorem 4.6 in other ways, e.g by considering the probability generating function (see section 4.4 below).

We now turn attention to higher moments, which, even in a TID model, will not be the same as classically because some of the sets of 3 edges we consider in the expression for $\mathbf{E}(\mathcal{E}^3)$ are cycles and so the probability that they arise will be different from classically (unless the model is trivial). Of course similar remarks apply to higher moments. However, we would expect (informally speaking) that most sets of r edges, where r is small compared to n, would contain few cycles, and so that, if the model is TID so that the probability of the other r-sets arising is the same as classically, the difference between $\mathbf{E}(\mathcal{E}^r)$ in the new and classical models would not be very large. To make these ideas more precise, if $\Gamma(n, k, \mathbf{s}, P)$ is TID the only sets of indicators which do not contribute the same amount to the new and classical moments are those which contain a cycle; thus if $f(n, r)$ is the probability that a graph with r edges chosen at random is a forest (that is, it has no cycles), then $\mathbf{E}(\mathcal{E}_\Gamma^r) - \mathbf{E}(\mathcal{E}_\alpha^r)$ will be a sum over the $(n(n-1)/2)^r(1 - f(n, r))$ cases where there are cycles present of some non-zero quantity; thus some insight into the difference between the two can be had by understanding $f(n, r)$.

Theorem 4.8 *Of the* $\binom{n(n-1)/2}{r}$ *total possible choices of* r *distinct edges on* n *labelled vertices,*

$$\sum_{k_1+\ldots k_{n-r}=n} \binom{n}{k_1, \ldots k_{n-r}} \prod_{i=1}^{n-r} k_i^{k_i-2}$$

are forests. If we choose r *edges, with replacement, from a set of* $n(n-1)/2$ *edges, with all choices being equiprobable, then the probability that the edges selected form a forest is*

$$f(n,r) = \sum_{s=0}^{r} \frac{\sum_{k_1+\ldots k_{n-s}=n\, k_i \geq 1} \binom{n}{k_1, \ldots k_{n-s}} \prod_{i=1}^{n-s} k_i^{k_i-2} \sum_{j=0}^{s} (-1)^j \binom{s}{j} (s-j)^r}{(n(n-1)/2)^r}.$$

Proof. A forest on n vertices with r edges has l, $1 \leq l \leq n$, components, which are trees. Thus if a component has m vertices it has $m-1$ edges. Thus if k_i, $1 \leq i \leq l$ are the numbers of vertices in the l components, we have $\sum_{i=1}^{l} k_i = n$ and $\sum_{i=1}^{l}(k_i - 1) = r$ so $l = n - r$. Since the number of trees on m labelled vertices is m^{m-2} by Cayley's theorem the first claim follows considering the various ways to select the vertex sets of the components and the tree within them.

For the second statement, we have, writing A_r for the event that r edges chosen randomly with replacement form a forest, we have

$$P\{A_r\} = \sum_{s=0}^{r} P\{A \mid s \text{ distinct edges }\} P\{\text{we get } s \text{ distinct edges}\}$$

$$= \sum_{s=0}^{r} P\{\text{s distinct edges form a forest}\} p_{n,r,s}$$

$$= \sum_{s=0}^{r} \frac{\sum_{k_1+\ldots k_{n-s}=n\, k_i \geq 1} \binom{n}{k_1,\ldots k_{n-s}} \prod_{i=1}^{n-s} k_i^{k_i-2}}{\binom{n(n-1)/2}{s}} p_{n,r,s}$$

by the previous paragraph, where $p_{n,r,s}$ is the probability that, in taking a sample of size r with replacement from a population of size $n(n-1)/2$, we get s different elements in our sample. This is clearly equal to the probability that in putting r balls into $n(n-1)/2$ cells with each ball equally likely to go in each cell so that there are $(n(n-1)/2)^r$ possible outcomes, we get $n(n-1)/2 - s$ empty cells, which by formula 11.7 on page 60 of [Fe] is

$$\frac{\binom{n(n-1)/2}{s} \sum_{j=0}^{s} (-1)^j \binom{s}{j} (s-j)^r}{(n(n-1)/2)^r}$$

and the result follows cancelling $\binom{n(n-1)/2}{s}$ top and bottom. •

This formula is of course rather intractable. For doing asymptotics, it may be worth noting that the number of ways to select r edges without replacement is $\binom{n(n-1)/2}{r}$ which provided r stays fixed as $n \to \infty$, is asymptotically the same as the total number of ways of selecting r edges with replacement namely $(n(n-1)/2)^r/r!$.

The above discussion also makes it clear that whether the first moment which is not the same in the new and classical models (the r-th moment, say) is greater than or less than classically will depend on whether r-cycles are more or less likely than classically; the techniques of Chapter 2 can be applied to this question.

If the model is not tree-indiscernible, even the second moment will differ from classically, considering two incident edges; by Theorem 4.7 it will be larger than classically.

To illustrate all this, consider $G_{p,q}$. Elementary calculations (which we in fact did on the computer using the generating function of $\mathcal{E}$) give

$$\mathbf{E}\left(\mathcal{E}_{p,q}^3\right) = \mathbf{E}\left(\mathcal{E}_{\alpha}^3\right) + n\,(n-1)\,(n-2)\left(\frac{p-q}{2}\right)^3$$

and thus the skewness $S = \mathbf{E}\left(((\mathcal{E}-\mu)/\sigma)^3\right)$ where μ and σ are the mean and standard deviation of $\mathcal{E}$, depends on p and q, rather than just their average; it is greater than classically if $p > q$ and less if $q > p$. Similarily to the calculation for skewness, we see that

$$\mathbf{E}\left(\mathcal{E}_{p,q}^4\right) - \mathbf{E}\left(\mathcal{E}_{\alpha}^4\right)$$

$$= \left(\frac{p-q}{2}\right)^3 \frac{n(n-1)(n-2)}{2} \left(q\left(2n+1\right)\left(n-3\right) + \left(2n+7\right)\left(n-3\right)p + 12\right)$$

so the kurtosis $K = \mathbf{E}((\mathcal{E}-\mu)/\sigma)^4)$ depends on both p and q rather than just their average; for $n \geq 3$, the right hand bracket is positive so the fourth moment is greater than classically if $p > q$ and less if $q > p$. We shall expand on these observations in section 4.5.

Note that in the above examples we have that (for $r = 3, 4$)

$$\lim_{n\to\infty} \frac{\mathbf{E}(\mathcal{E}_{p,q}^r)}{\mathbf{E}(\mathcal{E}_{\alpha}^r)} = 1 + O(n^{-3}).$$

If we allow r to vary with n, it is unrealistic to expect this to be true in general; for very large values of r, for example, most collections of r edges will contain several cycles, and so the difference will be greater. We do not know if some such statement does hold if we insist that r stay fixed.

4.4 Normal approximation of the number of edges

In the previous section we showed that in TID models $\mathbf{E}(\mathcal{E})$ and $\mathrm{Var}(\mathcal{E})$ depend only on α. One obvious question to ask next is whether there is asymptotic convergence of $\mathcal{E}$ to a normal distribution, if the p_{ij} do not depend on n; of course in the classical situation this is true by the De Moivre-Laplace theorem. We now show that it is also true here provided the model is not TID; we comment briefly on the situation when the model is TID below. In section 4.7 we shall discuss the situation where the probabilities are small for large n and where Poisson approximation is more appropriate.

The technique is Silverman's central limit theorem for exchangeable dissociated random variables. This requires some notation to state. Define an m-tuple to be an ordered set of m distinct positive integers $J = \{j_1, \ldots j_m\}$, and let $\mathcal{P}(m)$ be the set of all m-tuples and $\mathcal{P}(m,n)$ the set of all m-tuples whose elements are drawn from the set $\{1, 2, \ldots n\}$. (In our application $m = 2$

and the 2-tuples are the edges of the graph). For $J \in \mathcal{P}(m)$ and π a permutation of the integers, $J\pi$ is the m-tuple $(j_1\pi, ...j_m\pi)$, and a **rearrangement** of J is an m-tuple K with $\{j \in J\} = \{k \in K\}$.

$\mathcal{H}_{m\bullet}$ will denote an array of random variables X_J indexed by all $J \in \mathcal{P}(m)$. We will say that the array is **dissociated** if and only if the random variables in it are dissociated, in the sense of section 3.2.

A dissociated array is **exchangeably dissociated** if and only if for any finite sequence $J, ...Z$ of m-tuples and any permutation π of the integers, $(X_J, ...X_Z)$ and $(X_{J\pi}, ...X_{Z\pi})$ have the same distribution. Thus for $m = 2$ a dissociated array is exchangeably dissociated if and only if it is weakly exchangeable in the sense we defined in the discussion after Lemma 4.2.

Finally, note that $\mathcal{H}_{m\bullet}$ is **symmetrical** if and only if, whenever the m-tuples J and K are rearrangements of each other, we have $X_J = X_K$ a.s.; this will be true in our application as the edges are undirected. We can now state a weakened form of the theorem (it is easy to generalise the theorem to non-symmetrical exchangeably dissociated arrays, but we will not need this).

Theorem 4.9 *Suppose $\mathcal{H}_{m\bullet}$ is a symmetrical exchangeably dissociated array of real random variables with finite variance and zero mean. For each positive integer n let $S_n = \sum_{J \in \mathcal{P}(m,n)} X_J$. Then*

$$\frac{S_n}{n^{m-1/2}} \to N\left(0, m^2\rho\right)$$

in distribution as $n \to \infty$, where $\rho = Cov(X_J, X_K)$ for any two m-tuples J and K with exactly one element in common.

Proof. [Si]. •

Theorem 4.10 *In any RRC model which is not TID, $\mathcal{E}$ has an asymptotically normal distribution.*

Proof. We need only note that $S_n = 2(\mathcal{E} - n(n-1)\alpha/2)$ so that $\mathrm{Var}(S_n) = 4\mathrm{Var}(\mathcal{E})$. Thus the result follows from the fact that ρ above is non-zero because the model is not TID. •

If the model is TID, Silverman's theorem is of course still true, but as then $\rho = 0$ we have that this normalisation of $\mathcal{E}$ converges to a degenerate normal. A moment's thought shows that, since $\mathrm{Var}\mathcal{E}$ is the same as classically by Theorem 4.5, namely $n(n-1)\alpha(1-\alpha)/2$, if we want $(\mathcal{E}_n - n(n-1)\alpha/2)/(n^\kappa)$

to converge to a non-degenerate distribution, we must have $\kappa = 1$. We suspect that if the model is TID, this will converge, but do not have a proof at the moment.

It may well also be possible to imitate a more traditional method of proof of this fact, by showing that the moment generating functions $m_n(\theta)$ of $\mathcal{E}$ as n varies, suitably normalised in the usual way, converge to the moment generating function of a normal with mean 0 and variance 1. Indeed the generating function of $\mathcal{E}$ is easy to obtain;

Theorem 4.11 *In any RRC model $\Gamma(n,k,P,\mathbf{s})$, $\mathcal{E}_n$ has moment generating function $m_n(\theta)$ given by the following formula, with the sum taken over all integers $n_1,...n_k$ such that $n_i \geq 0$ for all i and $\sum_i n_i = n$;*

$$\sum \binom{n}{n_1,...n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{1\leq r<s\leq n} (p_{rs}e^{\theta} + 1 - p_{rs})^{n_r n_s} \prod_{r=1}^{k} (p_{rr}e^{\theta} + 1 - p_{rr})^{n_r(n_r-1)/2}.$$

Proof. This is a simple argument conditioning on the numbers of vertices of each colour and using the well-known fact that the moment generating function of a single Bernoulli trial with success probability α is $\alpha e^{\theta} + 1 - \alpha$. Indeed if we get n_i vertices of colour i (with probability $\binom{n}{n_1,n_2,...n_k} \prod_{j=1}^{k} s_j^{n_j}$) there are $n_i(n_i-1)/2$ potential edges between vertices of colour i, each of which arises independently with probability p_{ii} and $n_i n_j$ potential edges between vertices of colour i and colour j, which arise with probability p_{ij}. •

However, whilst it is probably possible to get the result this way, some of the details of convergence do not look altogether pretty. However, Theorem 4.11 will be useful in Chapter 5.

4.5 Stochastic dominance questions

In section 4.3 above, we obtained partial results on how close the moments of $\mathcal{E}$ are to their values in the classical model. A related question is to ask whether the moments are larger or smaller than their classical values. For the first moment which is not exactly equal to its classical value (the second moment if the model is not TID, and the third moment if the model is TID) we noted that the answer was straightforward; however, for higher moments, the answer is in general less clear; for example, in $G_{p,q}$ with $q > p$ the contribution from cycles of odd length will be less than classically by Theorem 2.6, but the contribution from 4-cycles will be more than classically,

and it is not obvious what the overall effect will be. However if we restrict attention to a smaller class of models, we can use techniques from Chapter 3 to attack the question. Our first result is the argument similar to Theorem 3.1 which was promised in the remarks after that theorem.

Theorem 4.12

$$\mathbf{E}(\mathcal{E}^r_{p,q}) \geq \mathbf{E}(\mathcal{E}^r_{\alpha})\forall r \ \textit{if} \ \ p > q.$$

Proof. We consider each of the $n(n-1)/2$ possible edges in the random graph on n vertices, letting K_n denote the set of all possible edges on these vertices. We take for each such edge a random variable $U_e \sim Un[0,1]$ and assume the various U_e are all independent as e varies. We describe the number of edges in the two models in terms of these variables and indicator variables I_e; two different ways of generating the I_e will give the two different models. Firstly, if we let $I_e \sim \text{Bin}(1,1/2)$, independently of each other and the U_e, then defining

$$\mathcal{E} = \sum_{e \in K_n} (I\{U_e < q\} + I_e I\{q \leq U_e < p\})$$

we see that the edge e contributes to the sum with probability q when $I_e = 0$ and with probability $p - q$ conditional on $I_e = 1$, so the overall probability that it contributes is $q + (p-q)/2 = (p+q)/2$, and the differing edges are independent, so the resultant random variable $\mathcal{E}$ is the number of edges in the G_α model.

On the other hand, we can define I_e to be 1 if and only if e is a same-same edge with respect to the random colouring of the vertices, and to be 0 otherwise; then as the edge in question contributes with probability q if I_e is zero, and with probability p if I_e is 1, we clearly get the $G_{p,q}$ model. (Note that this is effectively a **coupling** of the two distributions).

Now we work out $\mathbf{E}(\mathcal{E}^r)$ in both models. We will get a sum of expectations of products of r indicators (some of them possibly repeated) as in the inequality denoted $(*)$ in the proof of Theorem 3.1. Again as in that theorem, such a product is 1 if all the indicators involved are 1, and is 0 otherwise. As in Theorem 2.1, we note that if the I_e are defined by the first approach, the distinct I_e are independent, whereas in the second there may be correlation structure induced by cycles forcing an I_e to be 1 because the indicators of the other edges in the cycle are 1, so again if we can strip away edges to reduce to a maximal forest within the relevant subgraph, and see that the value of

the expectation in the case when the I_e are correlated with $p > q$ will be at least as large as when they are not correlated. •

Corollary 4.13 *If $p > q$, then $m_{\mathcal{E}_{p,q}}(\theta) \geq m_{\mathcal{E}_\alpha}(\theta)$ for $\theta > 0$.*

Proof. Apply theorem 4.12 recalling that

$$m_X(\theta) = \sum_{r=0}^{\infty} \frac{\mathbf{E}(X^r)\theta^r}{r!}. \bullet$$

Again it is natural to ask what happens if $q > p$. The analogue of the formula above is $\mathcal{E} = \sum_{e \in K_n} I\{U_e < p\} + I_e I\{U_e \in (p, q)\}$ but now to get the $G_{p,q}$ model we must say that $I_e = 1$ if and only if e is a red-blue edge, and so the nature of the correlation structure is more complex. As

$$\mathbf{E}\left(\mathcal{E}_{p,q}^3\right) = \mathbf{E}\left(\mathcal{E}_\alpha^3\right) + n\,(n-1)\,(n-2)\left(\frac{p-q}{2}\right)^3$$

we see that for $q > p$ there is a neighbourhood $\theta \in (-\epsilon, 0)$ in which $m_{\mathcal{E}_{p,q}}(\theta) \leq m_{\mathcal{E}_\alpha}$ by arguments similar to those used in Theorem 3.6.

It is natural to ask if we can generalise Theorem 4.12 to models other than $G_{p,q}$, for example in the same sort of way as Theorem 3.10 generalises Theorem 3.1, and this turns out to be true.

Theorem 4.14 *If $q < p < r$ we have*

$$\mathbf{E}(\mathcal{E}^r) \text{ in } G_{p,q,r} \geq \mathbf{E}(\mathcal{E}^r) \text{ in } G_\alpha \; \forall r.$$

Proof. This is similar to Theorem 4.12 and 3.10. This time, we have for each potential edge e two random variables U_e and V_e, where the U_e and V_e are all independent of each other, and all have a uniform distribution on $[0, 1]$. We also have for each edge two indicators I_e and J_e. We will show that under two regimes for generating the I_e and J_e the numbers of edges in $G_{p,q,r}$ and G_α are given by

$$\mathcal{E} = \sum_{e \in K_n} \left(I(U_e < q) + I_e I(q \leq U_e < p) + J_e I(p \leq U_e < r)\right) \quad (*)$$

and will use this to compare the moments in the two regimes. Firstly, we let

$$I_e = I(V_e < \frac{1}{2}) \sim \text{Bin}(1, \frac{1}{2}) \text{ and } J_e = I(V_e < \frac{1}{8} \text{ or } \frac{1}{2} \leq V_e \leq \frac{5}{8}) \sim \text{Bin}(1, \frac{1}{4})$$

so that the I_e and J_e are independent. Then in the formula $(*)$ we can only get one term on the right-hand side contributing for each edge as before, and a given edge does contribute with probability

$$q + \frac{p-q}{2} + \frac{r-p}{4} = \frac{p+2q+r}{4}$$

with the edges contributing or not independently of each other; thus this regime gives the model G_α.

Secondly, we consider the colouring in $G_{p,q,r}$ and let I_e be an indicator of whether or not the edge in question is same-same, and J_e be an indicator of whether the edge is blue-blue. Then again each edge contributes either 1 or 0 on the right-hand side of $(*)$. If the edge is red-blue, the contribution is 1 with probability q. If the edge is red-red, $I_e = 1$ and $J_e = 0$ so the edge contributes 1 with probability $q + (p - q) = p$. If the edge is blue-blue, it is automatically same-same, and so the probability that it contributes 1 is $q+(p-q)+(r-p) = r$. Hence under this regime we do generate the number of edges in $G_{p,q,r}$.

We now compare moments of $\mathcal{E}^r$ in the two regimes. As in the proof of Theorem 3.10, expanding out and considering the various terms being summed on each side, it is enough (as the indicators $I(U_e < q), I(q \leq U_e < p)$ and $I(p \leq U_e < r)$ are the same on both sides) to prove that the expectation of any product of I_es and J_es in the $G_{p,q,r}$ regime is at least as large as the corresponding product of I_es and J_es in G_α. The argument for this is the same as that in Theorem 3.10; after taking account of any extent to which a component of the I_e and the J_e intersect, and are thereby forced to be blue-blue rather than just same-same, the edges of a maximal forest in the G_α regime are independent, but on the other side the existence of cycles may force some indicators to be 1 automatically. •

As in Chapter 3, some further generalisation of the result may well be possible, but we have not investigated this.

4.6 Correlation structure and the degree sequence

Another possible manifestation of correlation structure is the extent of dependence of the degrees X_i of vertex i in some RRC model such as $G_{p,q}$. (Recall that the **degree** of a vertex in a graph is the number of vertices adjacent to that vertex). Of course in any RRC TID model, each X_i is a $\mathrm{Bin}(n-1, \alpha)$ random variable, but the degrees are dependent and so we

consider which aspects of that dependence structure manifest the differences between the two models. We first note that the **pairwise** correlation of the degrees is the same as classically.

Theorem 4.15 *The pairwise correlation of degrees in any TID RRC model is* $1/(n-1)$*; in particular, it does not depend on the model.*

Proof. The X_i satisfy $\sum_{i=1}^{n} X_i = 2\mathcal{E}$ and have the same distribution (being an exchangeable sequence) so we have (for $i \neq j$)

$$\mathbf{E}\left(\sum_{i=1}^{n} X_i\right)^2 = \mathbf{E}\left(4\mathcal{E}^2\right)$$

$$\Rightarrow n(n-1)\,\mathbf{E}(X_iX_j) + n\mathbf{E}\left(X_i^2\right) = 2n(n-1)\,\alpha(1-\alpha) + (n(n-1)\,\alpha)^2$$

$$\Rightarrow \mathbf{E}(X_iX_j) = \alpha(1-\alpha) + ((n-1)\,\alpha)^2 \text{ since } \mathbf{E}(X_i^2) = n\alpha(1-\alpha) + n^2\alpha^2.$$

$$\Rightarrow \operatorname{corr}(X_i, X_j) = \frac{\mathbf{E}(X_iX_j) - \mathbf{E}X_i\mathbf{E}X_j}{\sqrt{\operatorname{Var}(X_i)\operatorname{Var}(X_j)}} = \frac{1}{n-1}$$

as required. •

So we must look at more complex interactions to show up distinctions. Recall that the **degree sequence** of the graph is the set of degrees rearranged in non-increasing order, $d_1 \geq d_2 \geq \ldots. \geq d_n$; of particular interest are the minimal degree $\delta(G) = d_n$ and the maximum degree $\Delta(G) = d_1$. In the classical model, a great deal is known about the degree sequence; we summarise what we will need.

Theorem 4.16 *1. If* y *is a fixed real number with*

$$\lim_{n\to\infty} \frac{\alpha(1-\alpha)n}{(\log(n))^3} = \infty,$$

$$\lim_{n\to\infty} P\{\Delta < \alpha n + \sqrt{2\alpha(1-\alpha)\,n\log(n)}f(y,n)\} = e^{-e^{-y}}$$

where

$$f(y,n) = \left(1 - \frac{\log\log(n)}{4\log(n)} - \frac{\log\sqrt{2\pi}}{2\log(n)} + \frac{y}{2\log(n)}\right).$$

2. Suppose $m = m(n) = o(n)$ but $m(n) \to \infty$, and that $\omega(n) \to \infty$ arbitrarily slowly. Define x by $1 - \Phi(x) = m/n$, where Φ is the distribution function of a standard normal. Then a.e. G_α satisfies

$$| d_m - \alpha n - x\sqrt{\alpha(1-\alpha)n} | \leq \omega(n)\sqrt{\frac{\alpha(1-\alpha)n}{m\log(n/m)}}.$$

3. Suppose $m = o((\alpha(1-\alpha)n/\log(n))^{1/4})$, but $\lim_{n\to\infty} m(n) = \infty$. Then a.e. G_α is such that for any function $\omega(n)$ tending to zero as $n \to \infty$,

$$d_i - d_{i+1} \geq \sqrt{\alpha(1-\alpha)n/\log(n)}\omega(n)/m^2 \text{ if } i < m.$$

Also, if m and $\omega(n)$ go to infinity with n, a.e. G_α has

$$d_i - d_{i+1} \leq \frac{\omega(n)\sqrt{\alpha(1-\alpha)n}}{m^2\sqrt{\log(n)}} \text{ for some } 1 \leq i \leq m$$

so that the lower bound on the gaps is essentially best possible.

Proof. Part I is [B], Theorem III.3', page 61. Part 2 is [B] Theorem III.12, page 67. The two statements in part 3 are [B] Theorem III.15 (page 68) and Theorem III.16 (page 69) respectively. •

We now turn to our models. For the rest of this section, we work with $G_{p,q}$ with p and q independent of n; it may well be that some at least of our results work more generally, but we have not checked this. Initially, unsure what behaviour of Δ and δ to expect, we considered computer simulations of the maximum and minimum degrees, generating 25 graphs on 1000 vertices in $G_{p,q}$ with $p+q = 1$. These suggested (informally speaking) that the maximum and minimum degrees do not change much as the difference between p and q increases, until we reach a certain point (for our graphs, near to $p = 0.85$) where the maximum degree seems to become noticeably smaller and the minimum degree larger; that is there is a suggestion that the values become more tightly concentrated around the mean. This of course would imply, since the variance of $\mathcal{E}$ is the same in the two models by Theorem 4.6, that there must be more vertices of moderately high or moderately low degree in $G_{p,q}$ to compensate for the lack of vertices of very high or low degree.

We now try to turn these observations into mathematics. Note first that the minimum degree of a $G_{p,q}$ has the same distribution as the maximum

degree of a $G_{1-p,1-q}$ subtracted from $n-1$; hence, for the level of analysis we will carry out, we need only look at the maximum degree. We start with the easy observation that the distribution of Δ in $G_{p,q}$ with $\alpha \leq 1/2$ will differ from the distribution in G_α.

Lemma 4.17 *The distributions of the maximum degree in G_α and $G_{2\alpha,0}$ are different for any $\alpha \leq 1/2$.*

Proof. Suppose first $\alpha < 1/2$, so that both α and 2α satisfy the technical condition in Theorem 4.16 part 1. If $p = \alpha$, by Theorem 4.16, again recalling that

$$f(y,n) = \left(1 - \frac{\log\log(n)}{4\log(n)} - \frac{\log\sqrt{2\pi}}{2\log(n)} + \frac{y}{2\log(n)}\right)$$

we have

$$\lim_{n\to\infty} P\{\Delta < \alpha n + (2n\log(n)\alpha(1-\alpha))^{1/2} f(y,n)\} = e^{-e^{-y}}.$$

On the other hand, if $p = 2\alpha$ and $q = 0$ (possible by $\alpha \leq 1/2$), no red-blue edges arise, so the maximum degree is just the larger of the maximum degree of the reds and the maximum degree of the blues; letting m be the larger of the number of reds and the number of blues

$$\lim_{n\to\infty} P\{\Delta < 2\alpha m + (4\alpha(1-2\alpha)m\log(m))^{1/2} f(y,m)\} = e^{-e^{-y}}.$$

The new upper bound for Δ minus the old one has first few terms

$$2\alpha m + \sqrt{4\alpha(1-2\alpha)m\log(m)} - (\alpha n + (2n\log(n)\alpha(1-\alpha))^{1/2}).$$

Since $m = n/2 + cn^{1/2}$, where $c > 0$, this is dominated by the term

$$(n\log(n))^{1/2}(\sqrt{2\alpha(1-2\alpha)} - \sqrt{2\alpha(1-\alpha)})$$

which is negative unless $\alpha = 0$; thus the distribution of Δ differs from classically, with smaller values being more likely.

Next we consider what happens when $\alpha = 1/2$; then the technical condition in Theorem 4.16 no longer holds, so it does not give the behaviour of Δ when $p = 2\alpha$. However, if $p = 1$ and $q = 0$ the maximum degree is one less than the greater of the number of reds and the number of blues, so is,

for any function $\omega(n)$ which goes to infinity with n, $\leq n/2 + o(\sqrt{n}\omega(n))$ with probability tending to 1 as $n \to \infty$, including

$$\omega(n) = \left(\frac{\log(n)}{2}\right)^{\frac{1}{2}} \left(1 - \frac{\log\log(n)}{4\log(n)} - \frac{\log((2\pi)^{\frac{1}{2}}}{2\log(n)}\right)$$

so with overwhelming probability Δ is smaller than when $p = q = 1/2$. •

Of course, if $\alpha > 1/2$ we cannot consider $G_{2\alpha,0}$, but then in $G_{2\alpha-1,1}$

$$\Delta \simeq n/2 + n/2(2\alpha - 1) + \sqrt{2(2\alpha - 1)(2 - 2\alpha)n/2\log(n/2)}$$

$$\simeq n\alpha + \sqrt{2(2\alpha - 1)(1 - \alpha)n\log(n)}$$

which will be different from the classical maximum degree (unless $\alpha = 1$).

This much, which uses only a crude estimate of Δ in G_α shows that Δ does depend on both p and q rather than their average. However, we really want some kind of estimate for Δ in all $G_{p,q}$. We will get one by exploiting the more detailed results in Theorem 4.16. The proof will depend on understanding the maximum degree of a vertex in the graph of vertices its own colour and the maximum degree of a vertex in the bipartite graph of red-blue edges. We shall show that both these are tightly concentrated; and that the number of vertices of high degree in the graph of vertices their own colour is small, so that they do not have high degree in the bipartite graph also, whereas the vertices that do have high degree in the bipartite graph have low degree in the graph of vertices of their own colour.

The information on the maximal degree of a vertex in the graph of vertices of its own colour, and how many vertices have degrees close to Δ, was given in Theorem 4.16. We thus need to understand the maximum degree in a bipartite graph whose two vertex classes are a (judiciously chosen) subset of the red vertices, and the (about $n/2$) blue vertices, and each edge arises independently with probability q; in fact we will be considering the red vertices only, so need only consider, for each of the vertices of one colour, the number of edges to the opposite colour; thus the problem is to find the distribution of the maximum of some number $f(n)$ of **independent** random variables X_n, each of which is $\mathrm{Bin}(n, q)$. This question can be tackled using the following form of the DeMoivre-Laplace theorem;

Theorem 4.18 *Suppose* $0 < h = x\sqrt{q(1-q)n} = o((q(1-q)n)^{2/3})$. *Then, if* $Y_n \sim Bin(n, q)$, *we have* $P\{Y_n \geq qn + h\} \sim 1 - \Phi(x)$ *where* Φ *is the*

distribution function of the standard normal. In particular, if $x \to \infty$ *with* n

$$P\{Y_n \geq qn + h\} \sim \frac{e^{-x^2/2}}{\sqrt{2\pi}x}.$$

Proof. [B] Theorem I.6 •

We shall also use a form of the product formula for the exponential;

Lemma 4.19 *Let* a_n *be a sequence of real numbers. Then*

$$\lim_{n\to\infty} a_n = a \Rightarrow \lim_{n\to\infty}(1 - \frac{a_n}{n})^n = e^{-a}.$$

Conversely, if

$$\lim_{n\to\infty}(1 - \frac{a_n}{n})^n = c \in [0,1]$$

exists, then

$$c \in (0,1) \Rightarrow a_n \to a \in (0,\infty).$$

Proof. The first claim is standard, see [D] p112. For the rest, let $a = -\log_e(c) > 0$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty}(1 - \frac{a+\epsilon}{n})^n = e^{-a-\epsilon} < c$$

and so

$$(1 - \frac{a+\epsilon}{n})^n < (1 - \frac{a_n}{n})^n$$

for all large enough n, that is $a_n < a + \epsilon$. Similarily $a_n > a - \epsilon$ for all large enough n. •

This allows us to approximate the maximum degree in the bipartite graph.

Theorem 4.20 *The maximum of* $\lfloor n^\delta \rfloor$ *independent Bin(n, q) random variables (where* q *is constant and* $0 < \delta < 1$*) is* $qn + \sqrt{2\delta q(1-q)\log(n)n}$ *+smaller order terms.*

Proof. Note that, taking $x = \sqrt{k\log(n)}$ in Theorem 4.18, we have

$$P\{\max_{1\leq i\leq n^\delta} Y_i \leq qn + \sqrt{k\log(n)}\sqrt{q(1-q)n}\}$$

$$= P\{Y_1 \leq qn + \sqrt{k\log(n)}\sqrt{q(1-q)n}\}^{n^\delta}$$

$$= (1 - P\{Y_1 \geq qn + \sqrt{k\log(n)}\sqrt{q(1-q)n}\})^{n^\delta}$$

and so we must have, by Lemma 4.19, that

$$n^\delta P\{Y_1 \geq qn + \sqrt{k\log(n)}\sqrt{q(1-q)n}\} \to a$$

for some $a \in (0, \infty)$, if we are to get a non-degenerate distribution. But then n^δ times any quantity asymptotically equivalent to the probability must also tend to a; by Theorem 4.16, Theorem 4.18 and the fact that $\lim_{n\to\infty} x = \infty$, we must have

$$\lim_{n\to\infty} n^\delta \frac{e^{-k\log(n)/2}}{\sqrt{2\pi k\log(n)}} = n^\delta \frac{1}{\sqrt{2\pi k\log(n)}n^{k/2}} = a$$

and this expression will go to infinity as n goes to infinity if $k < 2\delta$ and to zero if $k > 2\delta$, giving the required result. •

We can now pull strings together. The idea will be to consider vertices in (say) the reds which are between the $n^{\delta-\epsilon}$th and $n^{\delta+\epsilon}$th in the degree sequence within the reds - we shall refer to these, for convenience, as these vertices. We will then consider the value of δ which maximises the implied upper bound $B + f(\epsilon)$ on the maximum degree amongst these vertices for small values of ϵ, using Theorems 4.16 and 4.20; and will then show that the actual maximum degree of these vertices is at least $B + g(\epsilon)$ where $0 < f - g \to 0$ as $\epsilon \to 0$; we will then show that the degree of the mth vertex in the degree sequence, where $m < n^{\delta-\epsilon}$ or $m > n^{\delta+\epsilon}$ cannot compete with the degree obtained above.

Theorem 4.21 *In $G_{p,q}$ a.e. graph has maximum degree (where dots denote higher order terms)*

$$\Delta = \alpha n + \sqrt{p(1-p) + q(1-q)}\sqrt{n\log(n)} + ...$$

Proof. We first note we can suppose that we have $n/2$ reds and $n/2$ blues, since the above arguments show that the slight variability in the number of reds and blues is unimportant. Again by the symmetry, we need only look at (say) the red vertices, since the top degree in the blues will be the same.

By Theorem 4.16 part 2, in a.e. G_p we have, if m is $o(n/2)$ but goes to infinity with $n/2$, we get (see the proof of Theorem III.12 in [B] for the assertion about x)

$$d_m - pn/2 \sim x\sqrt{p(1-p)n/2} \text{ where } x \sim \sqrt{2\log(n/2m)}$$

where d_m is the degree of the vertex in the subgraph of vertices of its own colour. Thus if $m = (n/2)^\delta$ for $1 > t > 0$, $x \sim \sqrt{2(1-\delta)\log(n/2)}$. Thus, for these vertices, x is about $\sqrt{2(1-\delta)\log(n/2)}$ and the maximum internal degree amongst these vertices is thus about $pn/2 + \sqrt{p(1-p)(1-\delta)n\log(n)}$.

Also, by Theorem 4.20, the probability that the maximum degree of these vertices in the red-blue graph is less than $qn/2 + \sqrt{2kq(1-q)(n/2)\log(n/2)}$ is, for large values of n about

$$(1 - \frac{e^{-k\log(n/2)}}{\sqrt{4\pi\delta\log(n/2)}})^{(n/2)^{\delta+\epsilon}-(n/2)^{\delta-\epsilon}}$$

$$= (1 - \frac{1}{\sqrt{4\pi\delta\log(n/2)}(n/2)^k})^{n/2^{\delta+\epsilon}-(n/2)^{\delta-\epsilon}}.$$

Since, for large enough values of n, $n^{\delta+\epsilon} - n^{\delta-\epsilon} > (1-\eta)n^{\delta+\epsilon}$ for any $\eta > 0$, by an obvious modification of Theorem 4.20 we see that this expression will go to 1 as $n \to \infty$ if $\delta + \epsilon < k$ and to 0 if $\delta + \epsilon > k$. Consequently, the maximum red-blue degree amongst these vertices is about (letting n go to infinity and ϵ to zero)

$$qn/2 + \sqrt{\delta q(1-q)n\log(n)}$$

hence the overall top degree amongst these vertices is bounded above by

$$\alpha n + (\sqrt{\delta q(1-q)} + \sqrt{(1-\delta)p(1-p)})\sqrt{n\log(n)}\,(*).$$

We choose the value d_{opt} of δ which maximises this expression; for this, we have

$$f(x) = \sqrt{x}\sqrt{q(1-q)} + \sqrt{(1-x)}\sqrt{p(1-p)}$$

$$\Rightarrow \frac{df}{dx} = x^{-1/2}\sqrt{q(1-q)} - (1-x)^{-1/2}\sqrt{p(1-p)}$$

so the unique turning point is when $(1-x)q(1-q) = xp(1-p)$, that is when

$$x = \frac{q(1-q)}{q(1-q) + p(1-p)}$$

and at this value of x, the second derivative of f is easily checked to be negative, so this is a maximum; and at this value of x we see easily that

$$f(x) = \sqrt{p(1-p) + q(1-q)}$$

(note this if $p = q$, $p = 2\alpha$ or $p = 0$ this upper bound agrees with the exact value in Theorem 4.16).

Next we show that this upper bound is attained (to within $\epsilon\sqrt{n\log(n)}$) by these vertices. For this, with $d_{opt} = q(1-q)/(q(1-q)+p(1-p))$, the internal degrees amongst these vertices are all about $pn/2+\sqrt{p(1-p)(1-d_{opt})n\log(n)}$ on letting ϵ go to zero. By the same variant on Theorem 4.20 as before, the maximum red-blue degree amongst these vertices is still (letting $\epsilon \to 0$) about

$$qn/2 + \sqrt{d_{opt}q(1-q)n\log(n)}$$

and so looking at whichever of these vertices has maximal red-blue degree, we do indeed get a vertex whose degree is within $\epsilon\sqrt{n\log(n)}$ of the upper bound $(*)$.

Finally we have to show that if $m < n^{d_{opt}-\epsilon}$, or $m > n^{d_{opt}+\epsilon}$, then the degree of that vertex is not large enough. For this, we note that the upper bound on the degrees in equation $(*)$ still applies; and so, as that upper bound has a unique maximum at $\delta = d_{opt}$, for such m the maximum degree will be less than the upper bound on the maximum degree which will be less than the value it attains near d_{opt}. •

Corollary 4.22 *The minimum degree is maximised when $p = \alpha$.*

Proof. It is a simple exercise in calculus to show that the function $p \to \sqrt{p(1-p)+(2\alpha-p)(1-2\alpha+p)}$ has its unique turning point at $p = \alpha$, which is a maximum by considering the second derivative. •

Our techniques do not yield as detailed an approximation to the maximum degree as in the classical model; this would need much more work. Note that the result makes it clear that the symmetry between p and q which we observed is genuine and not just an accident.

Theorem 4.21 also explains the observation that the maximum degree did not vary much until p or q took fairly extreme values, as the map $x \to x(1-x)$ is known to be fairly flat around $x = 1/2$; note however that this does depend on $p + q = 1$ so that $1/2$ is the average value.

Note that the next term in the estimate of the maximum degree is harder to get; it seems likely to be dominated by the variability in the numbers of reds and blues which is of order $\sqrt{n}$.

One might be tempted to believe that the vertex, v say, which is of top degree in the subgraph of vertices of whichever of the reds and the blues are

more numerous, might be likely to be the vertex of top degree in the whole graph if $p > q$. However it seems unlikely that this will be true in general. Indeed since the gaps between the degrees in Theorem 4.16 part 3 are only $o(n^{1/2})$ there seem to be a reasonable number of vertices whose degrees can overtake v when the bipartite degrees are added on.

Some more detailed statements which are possible about degrees and numbers of edges in the classical model will not hold here. For example, a result, whose importance for developing much of the theory is emphasised in [B] Chapter II.3 states (crudely speaking) that, for a wide range of values of α, most G_α have all vertices of about the same degree and all not too small subsets of the vertices have similar numbers of adjacent vertices.

Theorem 4.23 *1. Suppose* $0 < \alpha(n) < 1/2$. *Then for a.e.* G_α, *if* $U \subset V(G)$ *with* $\mid U \mid = u$, $u > 252log_e n$ *we have, writing* $E(U)$ *for the set of edges between the elements of* U

$$\mid E(U) - \frac{\alpha u(u-1)}{2} \mid \leq \left(\frac{7\alpha log_e n}{u}\right)^{\frac{1}{2}} \binom{u}{2}$$

2. Let $\epsilon \in \left(0, \frac{1}{6}\right)$, $\alpha(n) \leq \frac{1}{2}$. *Then for a.e.* G_α, $\forall W \subset V(G)$ *s.t* $\mid W \mid \geq \lceil \frac{6log_e n}{\epsilon^2 \alpha} \rceil$, *writing* $\Gamma(z)$ *for the set of vertices adjacent to* z*) we have*

$$\mid \{z \in V - W : \mid\mid \Gamma(z) \cap W \mid - \alpha \mid W \mid\mid \geq \epsilon\alpha \mid W \mid\} \mid \leq \frac{12log_e n}{\epsilon^2 \alpha}$$

Proof [B] page 44. •

Clearly these statements cannot be true in our models. Indeed, taking U first to be a large set of reds in a $G_{p,q,r}$ with $r \neq p$ and then a large set of blues (such sets exist with overwhelmingly high probability) the analogue of the first statement fails. To see that the analogue of the second statement fails consider $G_{p,q}$ with p large, q very small, W a large set of reds; then for z a blue vertex we have

$$\mid\mid \Gamma(z) \cap W \mid - \alpha \mid W \mid\mid \simeq \alpha \mid W \mid > \epsilon\alpha \mid W \mid$$

and in ${}_sG_{p,q,r}$ $(p \neq r)$ it is not even true that all vertices have about the same degree.

4.7 Poisson approximation and total variation distance

Earlier in this chapter we showed that, in the circumstances where the number of edges is approximated by a normal classically, this continues to be true in our models (unless they are TID). The other classical situation where one obtains a good approximation to a binomial is when the probability is small and we can use a Poisson approximation, and in this section we discuss how well we can do this for $\mathcal{E}_{p,q}$ and $\mathcal{E}_{\alpha}$, whose laws we denote by $\mathcal{L}_{p,q}$ and $\mathcal{L}_{\alpha}$ respectively. Our measure of closeness will be the following;

Definition 4.3 *The* **total variation distance** *between two probability laws taking values in* $\{0,1,2...\}$ *is*

$$d_{TV}(\lambda,\mu) = sup\{|\lambda(A) - \mu(A)| : A \subset N\} = \frac{1}{2}\sum_{j\geq 0} |\lambda\{j\} - \mu\{j\}|$$

Theorem 4.24 *Let* $(I_{i,j}, \{i,j\} \in E(G))$, *be a strongly dissociated family of* $Bin(1,\alpha)$ *random variables, indexed by the* N *edges of graph* G. *Let* $d_{i,j}$ *be the minimum of the degrees of* i *and* j, $W = \sum_{\{i,j\}\in E(G)} I_{i,j}$ *and* $\lambda = \mathbf{E}W = N\alpha$. *Then*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq (1-e^{-\lambda})2\alpha\left(\frac{\sum_{\{i,j\}\in\Gamma} d_{i,j}}{N} - \frac{1}{2}\right) \leq \frac{\sqrt{8}\left(1-e^{-\lambda}\right)\lambda}{N^{\frac{1}{2}}}.$$

In particular, if Γ *and* $(I_{i,j})$ *vary so that* $\lambda \to \lambda_{\infty}$ *and* $N \to \infty$, *keeping* $(I_{i,j})$ *strongly dissociated and equidistributed,* W *converges in distribution to a Poisson with mean* λ_{∞}.

If the $I_{i,j}$ *are merely dissociated rather than strongly dissociated, and* $\mathbf{E}(I_{i,j}I_{j,k}) = \alpha\sigma$ *for all* i,j *and* k, *we have*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{\left(1-e^{-\lambda}\right)}{\lambda}(N\alpha^2 + \sum_{\{i,j\}}(d_i + d_j - 2)(\alpha\sigma + \alpha^2))$$

$$= (1-e^{-\lambda})(\alpha + \frac{(\sigma+\alpha)\sum_{j=1}^{n} d_j(d_j-1)}{N}).$$

where d_i *is the degree of vertex* i *in the graph.*

Proof. The first paragraph is [BHJ] Theorem 2.O. The second is [BHJ] Corollary 2.N.1. •

[BHJ] gives an example (2.3.4) to show that the strong dissociation property is needed to get the full power of this result.

In our situation G is the complete graph on n vertices, so all vertices have degree $n-1$ and $N = n(n-1)/2$, and the variables are strongly dissociated if and only if the model is TID as remarked in Section 4.1. Thus we have the following corollary.

Corollary 4.25 *If we have a TID model, then if $\mathcal{L}$ is the law of the number of edges, then we have*

$$d_{TV}\left(\mathcal{L}\left(W\right), Po\left(\lambda\right)\right) \leq \left(1-e^{-\lambda}\right) 2\alpha \left(n - \frac{3}{2}\right)$$

so for any $\alpha = o\left(n^{-1}\right)$ we get a Poisson approximation result. For non-TID models, the result becomes

$$d_{TV}\left(\mathcal{L}\left(W\right), Po\left(\lambda\right)\right) \leq \frac{\left(1-e^{-\lambda}\right)}{\lambda}\left(\frac{n(n-1)\alpha^2}{2} + \sum_{\{i,j\}}(2n-4)(\alpha\sigma+\alpha^2)\right)$$

$$= (1-e^{-\lambda})(\alpha + (\sigma+\alpha)2(n-2)).$$

so that $\alpha = o(n^{-1})$ and $\sigma = o(n^{-1})$ suffices for convergence to a Poisson distribution, though with a less powerful constant.

Proof. Clear from the above. •

It is not clear whether we can improve significantly on this by exploiting additional structure in our indicators.

Note that the condition $\sigma = o(n^{-1})$ is in fact not needed; we must have $\max p_{ij} \sim \alpha$, and since $\mathbf{E}I_{i,j}I_{j,k} \leq \max\{p_{i,j}\}^2$ by a monotonicity argument, we have that $\sigma \leq \max\{p_{i,j}\}^2/\alpha$ and this is of the same asymptotic order as α, as required.

More generally we can ask about the total variation distance between the laws of $\mathcal{E}_{p,q}$ (say) and $\mathcal{E}_{\alpha}$. When Poisson approximation is sensible, which as we saw above is when p and q are $o(n^{-1})$ we have

$$d_{TV}\left(\mathcal{L}_{p,q}, \mathcal{L}_{\alpha}\right) \leq d_{TV}\left(\mathcal{L}_{p,q}, Po\left(\lambda\right)\right) + d_{TV}\left(Po\left(\lambda\right), \mathcal{L}_{\alpha}\right)$$

by the triangle inequality; Theorem 4.25 and the Prohorov estimate ([BHJ], page 2) of the distance between $\mathrm{Bin}(n,p)$ and $\mathrm{Po}(n\alpha)$ then yield an upper bound on the total variation distance, but that bound is likely to be very

weak. We can bound the total variation distance from below by the probability of any configuration of edges whose probability varies a lot from its value in the classical model; as different probabilities require cycles, and the probability of a triangle differs from its classical value most, namely by $((p-q)/2)^3$, we can consider $n/3$ independent triangles to get a lower bound $n((p-q)/2)^3/3$; again this is not very powerful.

5 Large deviations in the number of edges.

5.1 Probability theoretic background.

In this chapter, as promised in Chapter 3, we consider the probability of a large deviation in $\mathcal{E}$ the number of edges in our models. In general, if S_n is a sequence of random variables and $S_n = \sum_{i=1}^{n} X_i$, where the X_i are random variables each with mean μ, a large deviation principle is an estimate of $P\{S_n \geq n(\mu + \epsilon)\}$ for $\epsilon \neq 0$. Of course some kind of law of large numbers will usually ensure that this probability tends to 0 as n goes to infinity, but we aim to understand the rate at which it does so; in practice this usually means finding a suitable normalising function $f(n)$ for which

$$\lim_{n\to\infty} \frac{\log P\{\sum_{i=1}^{n} X_i \geq n(\mu + \epsilon)\}}{f(n)} = c,$$

where $c \neq 0$ is a constant; this is (unsurprisingly) more difficult.

The relevance of this to random graph theory is that there we often want to show that if some result leads to something interesting happening with small failure probability, then if we apply the result exponentially many times, the failure probability remains small (note that a polynomially small error probability could still blow up if we applied the result exponentially many times). So, for example, historically, it was not possible to improve the estimate that for a.e. G_α $(0 < \alpha < 1)$ the chromatic number $\chi(G_\alpha)$ is at most $(1 + o(1))n/(2\log_d n)$ (here $d = 1/(1-\alpha)$) to a proof that this bound is the correct asymptotic value, until it had been proven that the probability that the independence number of a graph is unexpectedly small is exponentially small; see [B1], Chapter 4 section 1 for more details of all this.

The basic case of large deviation theory is when the X_i are i.i.d. It will be helpful to recall the basic theorem, to illustrate how the moment generating function relates to the large deviation probability;

Theorem 5.1 *Let X_i be i.i.d with mean 0. Suppose the moment generating function $M(t)$ of X_i is finite in some interval around 0. Then if $a > 0$ and $P\{X_i > a\} > 0$,*

$$\lim_{n\to\infty} P\{\sum_{i=1}^{n} X_i > na\}^{1/n} = \inf_{t>0} \left(e^{-at} M(t)\right)$$

where the expression on the right-hand side is $\in (0,1)$.

Proof. This is standard; see e.g [GS], Theorem 5.11 •

We note that some work has already been done on generalisation of the upper bound (which is usually the easier part of a large deviations principle to prove) to the case where the X_i are not identically distributed, but are still independent; the following theorem is an example.

Theorem 5.2 *Let $X_1, \ldots X_n$ be independent Bernoulli rv's with $\mathbf{E}(X_i) = p_i$, $P = \sum_{i=1}^n p_i/n$ and $Q = 1 - P$. Then, for all $0 < t < Q$, we have*

$$P\{\sum_{i=1}^{n} X_i \geq n\,(P+t)\} \leq \left(\left(\frac{P}{P+t}\right)^{P+t}\left(\frac{Q}{Q-t}\right)^{Q-t}\right)^n.$$

Proof See [McD]. •

In our models, if the X_i $1 \leq i \leq n(n-1)/2$ are (dependent) indicators of whether or not each edge is present, Theorem 5.1 and a monotonicity argument show that, with $p_{\max} = \max_{i,j} p_{ij}$,

$$P\{\sum_{j=1}^{n(n-1)/2} X_j \geq \frac{(p_{\max}+\epsilon)n(n-1)}{2}\}$$

is exponentially small in $n(n-1)/2$, and putting $p_{\min} = \min_{i,j} p_{ij}$, there is of course a similar result for $P\{\sum_{j=1}^{n(n-1)/2} X_j \leq (p_{\min} - \epsilon)n(n-1)/2\}$. However, quite apart from the fact that this does not give the exact rate of decay, it does not tell us anything about what happens for values between $p_{\min}$ and $p_{\max}$.

Thus we want a large deviation principle which allows for some degree of dependence amongst the X_i. An appropriate framework for the kind of problems we will study here is that of the so-called Gartner-Ellis theorem. (Not all large deviation principles can be obtained in this way; for example [O] gives a large deviation principle for the order of the giant component of a G_α where the rate function (see below) is not convex so cannot arise from the Gartner-Ellis theorem). General discussion of Gartner-Ellis theory can be found in for example [DZ]. Let us merely note that such theorems are in general quite delicate and some care is needed in stating the exact form of the theorem needed for a particular situation. The particular form we use here, which is designed to cope with the difficulties arising when the rate function is not differentiable at the origin, was first used by Biggins and Bingham [BB]. Since their proof was omitted there, we give it here.

Before the proof proper, it will be helpful to make a few observations about convex functions. Suppose ϕ is a convex function which is differentiable at θ, which is in the interior of the set where ϕ is finite, and that $y = \phi'(\theta)$. Then, defining

$$I(y) = \sup_{\mu}(y\mu - \phi(\mu))$$

we have

$$I(y) = y\theta - \phi(\theta)$$

and for any $\epsilon > 0$

$$I(y+\epsilon) - I(y) - \theta\epsilon > 0.$$

This is almost obvious; in detail, note first that, as ϕ is differentiable at θ, it is finite in some neighbourhood of θ. Clearly $I(y) \geq y\theta - \phi(\theta)$. If we had $I(y) > y\theta - \phi(\theta)$, then, by the definition of supremum, there would exist a μ with

$$y\mu - \phi(\mu) > y\theta - \phi(\theta) \Rightarrow y(\mu - \theta) > \phi(\mu) - \phi(\theta).$$

But, since ϕ is convex, we know that $\eta(\mu) = (\phi(\mu) - \phi(\theta))/(\mu - \theta)$ is a nondecreasing function of μ by [We, 5.1.2]; hence in particular, if $\mu > \theta$, we have

$$y > \lim_{x \to \theta} \frac{\phi(x) - \phi(\theta)}{x - \theta} = y$$

giving a contradiction; and a similar argument deals with the case when $\mu < \theta$. Thus $I(y) = y\theta - \phi(\theta)$ as required. For the second claim,

$$I(y+\epsilon) - I(y) = \sup_{\mu}(y+\epsilon)\mu - \phi(\mu) - y\theta + \phi(\theta)$$

$$\geq (y+\epsilon)\theta - \phi(\theta) - y\theta + \phi(\theta) = \epsilon\theta.$$

It remains to show that we cannot have equality. If we did, we would have

$$\begin{aligned}(y+\epsilon)\mu - \phi(\mu) \leq (y+\epsilon)\theta - \phi(\theta)\,\forall\mu \quad &\Rightarrow \quad \phi(\mu) - \phi(\theta) \geq (y+\epsilon)(\mu - \theta) \\ &\Rightarrow \quad \frac{\phi(\mu) - \phi(\theta)}{\mu - \theta} \geq y + \epsilon\,\forall\mu > \theta\end{aligned}$$

Hence, if $\epsilon > 0$, this would imply

$$y = \lim_{\mu \to \theta} \frac{\phi(\mu) - \phi(\theta)}{\mu - \theta} \geq y + \epsilon$$

giving the desired contradiction. A similar argument, considering $\mu < \theta$ deals with the case $\epsilon < 0$. We now proceed to the theorem proper.

Theorem 5.3 *Suppose we have a sequence of random variables S_n and a_n, a sequence of positive numbers tending to infinity. Define*

$$\phi_n(\theta) = \frac{\log(\mathbf{E}(e^{\theta S_n}))}{a_n}$$

We assume that

$$\lim_{n\to\infty} \phi_n(\theta) = \phi(\theta)$$

exists pointwise (we allow ϕ and the ϕ_n to be infinite; note it is the assumption of the existence of this limit, in analogy to what would happen if the S_n were sums of i.i.d variables, when of course $\phi_n = \phi$ for all n, that limits the dependence. Note also that ϕ, being a limit of convex functions, is convex). We define the **rate function** *$I(y)$ by*

$$I(y) = \sup_{\mu} (\mu y - \phi(\mu)).$$

For any function ϕ, define $D_\phi = \{x : \phi(x) < \infty\}$ and when $Int(D_\phi)$ is non-empty, define ϕ'_+ and ϕ'_- to be the right and left derivatives of ϕ (which exist as ϕ is convex. Since ϕ is convex so is D_ϕ). Then if $\exists s > 0 \in D_\phi$, we have

$$\liminf_{n\to\infty} \frac{-\log(P\{S_n \geq a_n x\})}{a_n} \geq I(x) \ \forall x > \phi'_+(0).$$

If in addition $y = \phi'(\theta)$ for some $\theta \in Int(D_\phi)$, so ϕ is differentiable at θ, then

$$\limsup_{n\to\infty} \frac{-\log(P\{S_n > a_n x\})}{a_n} \leq I(y) \ \forall x < y.$$

Proof. By Markov's inequality,

$$P\{S_n \geq a_n y\} e^{\theta a_n y} \leq \mathbf{E} e^{\theta S_n} = e^{a_n \phi_n(\theta)}$$

provided $\theta > 0$; hence

$$\liminf_{n\to\infty} \frac{-\log P\{S_n \geq a_n y\}}{a_n} \geq \sup_{\mu>0} (y\mu - \phi(\mu)).$$

Thus if $y > \phi'_+(0)$ the concave function $y\mu - \phi(\mu)$ is increasing at $\mu = 0$ and so the final bound is indeed $I(y)$.

In the other direction, if ζ_n is the measure corresponding to S_n we define the **conjugate probability measure**

$$\zeta_n^\theta(du) = e^{\theta u - a_n \phi_n(\theta)} P\{S_n \in du\},$$

with associated random variables S_n^θ. Then, denoting the ball of radius ϵ around a point y by $B_\epsilon(y)$, we have for small ϵ that

$$\begin{aligned} P\{S_n > a_n(y-\epsilon)\} &\geq P\{S_n \in a_n B_\epsilon(y)\} \\ &= e^{a_n\phi_n(\theta)-\theta y a_n} \int_{a_n B_\epsilon(y)} e^{\theta y a_n - \theta u} \zeta_n^\theta(du) \\ &\geq e^{(a_n\phi_n(\theta)-\theta y a_n)} e^{-\theta\epsilon a_n} P\{S_n^\theta \in a_n B_\epsilon(y)\}. \end{aligned}$$

As $I(y) = y\theta - \phi(\theta)$, the proof will be completed (taking logarithms, letting n go to infinity and finally letting ϵ go to zero) if we can show that

$$P\{S_n^\theta \in a_n B_\epsilon(y)\}$$

converges to 1 sufficiently quickly.

For this, note that we have

$$\lim_{n\to\infty} \frac{\log(\mathbf{E}(e^{sS_n^\theta})}{a_n} = \phi_n(s+\theta) - \phi_n(\theta)$$

(s being the dummy variable in the generating function). Since by assumption $\theta \in \mathrm{Int}(D_\phi)$, and the derivative of $\phi(s+\theta) - \phi(\theta)$ with respect to s at $s = 0$ is y by the remarks before the theorem, we can apply the first part of the theorem to the random variables S_n^θ to obtain that

$$\begin{aligned} \liminf_{n\to\infty} \frac{-\log P\{S_n^\theta \geq a_n(y+\epsilon)\}}{a_n} &\geq I(y+\epsilon) + \phi(\theta) - \theta(y+\epsilon) \\ &= I(y+\epsilon) - I(y) - \theta\epsilon. \end{aligned}$$

Since $I(y)$ is a convex function, with derivative θ at y, the remarks before this theorem imply this bound is positive, and so

$$P\{S_n^\theta \geq a_n(y+\epsilon)\}$$

goes to zero like the sequence $e^{-a_n\delta}$ for suitable $\delta > 0$. The same argument applied to the sequence $-S_n^\theta$ shows that

$$P\{S_n^\theta \leq a_n(y-\epsilon)\}$$

goes to zero like $e^{-a_n\gamma}$ for suitable $\gamma > 0$. Hence

$$0 \geq \frac{\log P\{S_n^\theta \in a_n B_\epsilon(y)\}}{a_n} \geq \frac{\log\left(1 - 2e^{-a_n\delta}\right)}{a_n} \to 0$$

completing the proof. •

An obvious irritation here is that there are two separate statements for the liminf and the limsup, when we would prefer to have a statement about the limit. In general, this cannot be guaranteed, but mild assumptions will give such a statement.

Corollary 5.4 *Suppose* $\exists s > 0 \in D_\phi$. *Then,* $\forall y \in Int\{y : y = \phi'(\mu)\}$ *for some* μ, *we have*

$$\lim_{n\to\infty} \frac{\log\left(P\{S_n \geq a_n y\}\right)}{a_n} = I(y).$$

Proof. This is an immediate corollary of the previous theorem, since in the interior of the region where ϕ is finite, the rate function is continuous; see e.g [BB] Corollary 1. •

The link with our situation of course arises by taking S_n to be the sum of the $n(n-1)/2$ X_j so that we must take $a_n = n(n-1)/2$.

5.2 A large deviations principle for $\mathcal{E}_{p,q}$.

In this section we will prove a large deviations result for the model $G_{p,q}$ giving full information on what happens in that model. In the following section we shall give a partial result valid for any RRC model.

The first step, of course, is to obtain the function ϕ in the statement of Theorem 5.3. A key role in our derivation will be played by the following lemma, which will occur again in later chapters.

Lemma 5.5 *For* $0 < x < 1$,

$$\lim_{n\to\infty} \frac{2\log\left(2^{-n}\sum_{i=0}^{n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)} = \log(1+x)$$

with the error term being $O\left(\frac{1}{n}\right)$. *For* $-1 < x < 0$,

$$\lim_{n\to\infty} \frac{2\log\left(2^{-n}\sum_{i=0}^{n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)} = \frac{\log(1-x^2)}{2}$$

with the error term again being $O\left(\frac{1}{n}\right)$. *For* $x = 0$, *the expression (and so its limit) is* 0.

Proof. The statement for $x = 0$ is trivial. Suppose $x > 0$. We have

$$\frac{2\log(2^{-n}\sum_{i=0}^{n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)})}{n(n-1)}$$

$$= \frac{2\log\left(2^{-n}(1+x)^{(n^2-n)/2}\sum_{i=0}^{n}\binom{n}{i}\left(\frac{1-x}{1+x}\right)^{i(n-i)}\right)}{n(n-1)}$$

$$= \log(1+x) + \frac{2\log\left(\sum_{i=0}^{n}2^{-n}\binom{n}{i}\left(\frac{1-x}{1+x}\right)^{i(n-i)}\right)}{n(n-1)}$$

which, since $0 < (1-x)/(1+x) < 1$, is

$$\leq \log(1+x) + \frac{2\log\left(\sum_{i=0}^{n}2^{-n}\binom{n}{i}\right)}{n(n-1)} = \log(1+x)$$

since of course $\sum_{i=0}^{n}2^{-n}\binom{n}{i} = 1$. This gives an upper bound; to get a lower bound, we note that, just considering the $i = 0$ term, we have

$$\frac{2\log\left(\sum_{i=0}^{n}2^{-n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)}$$

$$\geq \frac{2\log\left(2^{-n}\binom{n}{0}(1+x)^{(n^2-n)/2}\right)}{n(n-1)}$$

$$= \log(1+x) - \frac{2n\log(2)}{n(n-1)} = \log(1+x) + O\left(\frac{1}{n}\right)$$

as $n \to \infty$, and the result follows.

Finally if $-1 < x < 0$ so that $(1-x)/(1+x) > 1$ we have

$$\frac{2\log\left(\sum_{i=0}^{n} 2^{-n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)}$$

$$\leq \frac{2\log\left((1+x)^{(n^2-n)/2}\sum_{i=0}^{n} 2^{-n}\binom{n}{i}\left(\frac{1-x}{1+x}\right)^{\lfloor\frac{n^2}{4}\rfloor}\right)}{n(n-1)}$$

since the function $i(n-i)$ is maximised as i ranges over $(0,n)$ when i is as close as possible to $n/2$. The above expression is asymptotically

$$\log(1+x)+\frac{n\log(\frac{1-x}{1+x})}{2(n-1)} = \log(1+x)+\frac{1}{2}\log(\frac{1-x}{1+x})+O\left(\frac{1}{n}\right) = \frac{\log(1-x^2)}{2}+O(\frac{1}{n}).$$

This gives the upper bound; to obtain a lower bound, this time, rather than looking at the case when i is as small or large as possible, we look at the term when $i = \lfloor n/2 \rfloor$, observing that

$$\frac{2\log\left(\sum_{i=0}^{n} 2^{-n}\binom{n}{i}(1+x)^{(n^2-n)/2-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)}$$

$$\geq \frac{2\log\left((1+x)^{(n^2-n)/2}2^{-n}\binom{n}{\lfloor n/2\rfloor}\left(\frac{1-x}{1+x}\right)^{\lfloor n/2\rfloor(n-\lfloor n/2\rfloor)}\right)}{n(n-1)}.$$

Now by Stirling's formula,

$$\frac{\binom{n}{\lfloor\frac{n}{2}\rfloor}}{2^n} \sim \frac{1}{\sqrt{2\pi n}} \geq \frac{1}{\sqrt{7n}}$$

for sufficiently large n, so then the above is

$$\log(1+x)+\frac{n\log\left(\frac{1-x}{1+x}\right)}{2(n-1)}+\frac{2\log\left(\frac{1}{\sqrt{7n}}\right)}{n(n-1)}$$

$$= \log(1+x)+\frac{1}{2}\log(\frac{1-x}{1+x})+O(\frac{1}{n}) = \frac{\log(1-x^2)}{2}+O(\frac{1}{n})$$

as n goes to infinity, so that the result again follows. •

It is worth noting that the lemma uses estimates which are in some sense very crude; in both cases, note that one of the bounds is obtained by ignoring all but one term in the summation, however which term varies (dramatically) between the two cases. We shall comment more on this later.

We can now obtain the function $\phi(\theta)$ in the Gartner-Ellis theorem;

Theorem 5.6

$$\phi(\theta) = \log\left(pe^{\theta} + 1 - p\right) \text{ if } \frac{(p-\alpha)(e^{\theta}-1)}{\alpha(e^{\theta}-1)+1} > 0,$$

$$\phi(\theta) = \frac{\log\left(pe^{\theta} + 1 - p\right) + \log\left(qe^{\theta} + 1 - q\right)}{2} \text{ if } \frac{(p-\alpha)(e^{\theta}-1)}{\alpha(e^{\theta}-1)+1} < 0$$

and is 0 otherwise.

Proof. We have

$$\phi_n(\theta) = \frac{2\log\left(\mathbf{E}\left(e^{\theta S_n}\right)\right)}{n(n-1)}$$

$$= \frac{2\log\left(\sum_{i=0}^{n} 2^{-n}\binom{n}{i}\left(pe^{\theta}+1-p\right)^{(n^2-n)/2-i(n-i)}\left(qe^{\theta}+1-q\right)^{i(n-i)}\right)}{n(n-1)}$$

$$= \log\left(\alpha e^{\theta}+1-\alpha\right) + \frac{\log\left(\sum_{i=0}^{n} 2^{-n}\binom{n}{i}(1+x)^{\frac{n^2-n}{2}-i(n-i)}(1-x)^{i(n-i)}\right)}{n(n-1)}$$

where $x = ((p-\alpha)(e^{\theta}-1))/(\alpha(e^{\theta}-1)+1)$; we use here the fact that $(pe^{\theta}+1-p)/(\alpha e^{\theta}+1-\alpha) = 1+x$ and that $(qe^{\theta}+1-q)/(\alpha e^{\theta}+1-\alpha) = 1-x$ which is easily checked from the definitions. Also, as $1+x$ and $1-x$ are both positive, we can apply Lemma 5.5, together with the fact that

$$\log\left(\alpha e^{\theta}+1-\alpha\right) + \log\left(1 + \frac{(p-\alpha)(e^{\theta}-1)}{\alpha e^{\theta}+1-\alpha}\right) = \log\left(pe^{\theta}+1-p\right)$$

to obtain the first statement of the theorem. For the rest, we have

$$\log\left(\alpha e^{\theta}+1-\alpha\right) + \frac{1}{2}\log\left(1 - \left(\frac{(p-\alpha)(e^{\theta}-1)}{\alpha(e^{\theta}-1)}\right)^2\right)$$

$$= \log\left(\alpha e^{\theta} + 1 - \alpha\right) + \frac{1}{2}\log\left(\frac{(\alpha(e^{\theta}-1)+1)^2 - ((p-\alpha)(e^{\theta}-1))^2}{(\alpha(e^{\theta}-1)+1)^2}\right)$$

$$= \frac{\log(p(e^{\theta}-1)+1) + \log(q(e^{\theta}-1)+1)}{2}$$

on factoring the difference of squares on the top of the right-hand side and simplifying.

Finally, the statement when $x = 0$ is clear. •

Corollary 5.7 *If* $p > \alpha$,

$$\phi(\theta) = \log\left(pe^{\theta} + 1 - p\right) \text{ if } \theta > 0,\ 0 \text{ if } \theta = 0 \text{ and}$$

$$\frac{\log(pe^{\theta} + 1 - p) + \log(qe^{\theta} + 1 - q)}{2} \text{ if } \theta < 0.$$

If $p < \alpha$,

$$\phi(\theta) = \log\left(pe^{\theta} + 1 - p\right) \text{ if } \theta < 0,\ 0 \text{ if } \theta = 0 \text{ and}$$

$$\frac{\log(pe^{\theta} + 1 - p) + \log(qe^{\theta} + 1 - q)}{2} \text{ if } \theta > 0.$$

Proof. This is immediate from the previous theorem (5.6) on unwinding the condition on $x = (p-\alpha)(e^{\theta}-1)/(\alpha(e^{\theta}-1)+1)$ into one on θ. •

Note that this result is compatible with Corollary 4.13 showing that when $p > q$ we have $m_{\mathcal{E}_{p,q}}(t) \geq m_{\mathcal{E}_{\alpha}}(t)$ for $t \geq 0$ and that $\exists\, \epsilon > 0$ such that $m_{\mathcal{E}_{p,q}}(t) \leq m_{\mathcal{E}_{\alpha}}(t)$ for $t \in (-\epsilon, 0)$.

Given ϕ, it is easy to complete the derivation of the rate function;

Corollary 5.8 *When* $p > q$, *the rate function* $I(y)$ *is the maximum of*

$$\sup_{\theta>0}\left(\theta y - \log\left(pe^{\theta} + 1 - p\right)\right),\ 0 \text{ and}$$

$$\sup_{\theta<0}\left(\theta y - \frac{\log(pe^{\theta} + 1 - p) + \log(qe^{\theta} + 1 - q)}{2}\right).$$

When $p < q$, *the rate function is the maximum of*

$$\sup_{\theta<0}\left(\theta y - \log\left(pe^{\theta} + 1 - p\right)\right),\ 0 \text{ and}$$

$$\sup_{\theta>0}\left(\theta y - \frac{\log(pe^{\theta} + 1 - p) + \log(qe^{\theta} + 1 - q)}{2}\right).$$

Proof. This is immediate from the previous corollary. •

Thus different rules govern large deviations above and below the mean. Note that if $y > p > \alpha$ the rate function is the same as in independent Bernoulli trials with probability p, as if $p > q$ so that $pe^{\theta} + 1 - p < qe^{\theta} + 1 - q$ for $\theta < 0$, we have

$$\sup_{\theta<0}\left(\theta y - \frac{\log(pe^{\theta}+1-p)+\log(qe^{\theta}+1-q)}{2}\right) \leq \sup_{\theta<0}\left(\theta y - \log\left(\alpha e^{\theta}+1-\alpha\right)\right).$$

A similar argument will of course work when $y < p < \alpha$.

In the classical case, when the rate function is

$$I(y) = \sup_{\theta}\left(\theta y - \log\left(\alpha e^{\theta}+1-\alpha\right)\right)$$

it is easy to show by elementary calculus that the supremum on the right-hand side is attained for $\theta = \log(y(1-\alpha)/((1-y)\alpha)$. For the range of values for which the rate function is an average of the rate functions for p and q we can in principle differentiate, solve a quadratic equation in e^{θ} and take logarithms to find the value of θ at which the function has a turning point (which is a maximum by concavity of the rate function). Unfortunately, the expression obtained is very intractable. Calculations suggest the value of θ for which the rate function is maximised is slightly greater than classically if $y > \alpha$ and smaller if $y < \alpha$ but even this does not seem easy to prove.

The relationship between this analysis and the question of how close the moments of the number of edges in the new model and the classical model are merits brief comment. Recall that we have shown that $\phi(\theta) \neq \log\left(\alpha e^{\theta}+1-\alpha\right)$ although they would have been equal if there had not been the correlation structure. Since the left-hand side is the limit of the normalised cumulant generating functions of $\mathcal{E}$ in our model, and the right-hand side is the limit of the normalised cumulant generating function in the classical model, and the cumulant generating function is well-known to be closely linked with the moments of the distribution, one might naively imagine that the fact that the two limits are different would mean that the moments of the two distributions $\mathcal{E}_{p,q}$ and $\mathcal{E}_{\alpha}$ cannot be very close to each other. However the link with the moments depends on differentiability of the functions at θ=0, and since we have seen that this does not hold for the limiting function, the situation is not too clear.

There is a subtlety which produces some complications;

Lemma 5.9 *If $p > q$ $\phi(h)$ is differentiable $\forall h \neq 0$ and takes all values in $(p,1)$ and $\left(0, \frac{p+q}{2}\right)$, but is not differentiable at $h = 0$.*

If $q > p$ $\phi(h)$ is differentiable $\forall h \neq 0$ and takes all values in $\left(\frac{p+q}{2}, 1\right)$ and $(0,p)$, but is not differentiable at $h = 0$.

Proof. Only the statement about behaviour at $h = 0$ requires proof. Assume $p > \alpha$. At $h = 0$ we have

$$\lim_{h\to 0+}\left(\frac{\phi(h)-0}{h}\right) = \lim_{h\to 0+}\left(\frac{\log(p(e^h-1)+1)}{h}\right)$$

$$= \lim_{h\to 0+}\left(\frac{p(e^h-1)}{h} - \frac{(p(e^h-1))^2}{h} + \ldots\right) = p$$

and by an analogous argument

$$\lim_{h\to 0-}\left(\frac{\phi(h)-0}{h}\right) = \frac{(p+q)}{2},$$

proving the claim. The argument for $p < \alpha$ is similar. •

Thus, whilst the above result coupled with Corollary 5.8 shows that we get a meaningful estimate of

$$\lim_{n\to\infty} \frac{-\log(P\{S_n \geq \frac{n(n-1)}{2}x\})}{\frac{n(n-1)}{2}}$$

for $x \in (p,1)$ and $\left(0, \frac{p+q}{2}\right)$ if $p > q$, and for $x \in \left(\frac{p+q}{2}, 1\right)$ and $(0,p)$ if $q > p$, we do not get any meaningful lower estimate of the probability for values of x between p and $(p+q)/2$; the above limit will just be degenerate.

This raises the question of whether there is a real rate function for the missing gap; for example, it might be the case that in that interval

$$\lim_{n\to\infty} \frac{\log(\mathbf{E}(e^{\mathcal{E}_n}))}{n^\kappa}$$

exists and is non-trivial for some choice of $0 < \kappa < 2$. The right choice turns out to be $\kappa = 1$, as the following theorem shows.

Theorem 5.10 *Suppose c is between p and α. Then*

$$\lim_{n\to\infty} \frac{\log\left(P\{\mathcal{E}_n \geq \frac{cn(n-1)}{2}\}\right)}{n} = \mu(c)$$

where $\mu(c) = c\log(c) + (1-c)\log(1-c) + \log(2)$

Proof. There are three disjoint ways in which such a large deviation can arise;

1. There is a large deviation in the number N_1 of reds, and then a number of edges asymptotically equivalent to the number we would expect to arise, conditional on the number of reds, do arise; then $N_1 \sim (1/2 + \epsilon)n$ for some $\epsilon \neq 0$ and the number of edges is of order of magnitude

$$p\left(\frac{N_1(N_1-1)}{2} + \frac{(n-N_1)(n-N_1-1)}{2}\right) + qN_1(n-N_1).$$

2. There is a large deviation in the number of reds, and then there is a large deviation in the number of edges arising, which however is **not** asymptotically the number we would expect given the large deviation in the number of reds; then $N_1 \sim (1/2 + \epsilon)n$ for some $\epsilon \neq 0$ but the difference between the number of edges and

$$p\left(\frac{N_1(N_1-1)}{2} + \frac{(n-N_1)(n-N_1-1)}{2}\right) + qN_1(n-N_1)$$

is of order of magnitude cn^2 for a suitable constant

3. There is no large deviation in the number of reds, but there is a large deviation in the number of edges arising; then $N_1 \sim n/2$ but $\mathcal{E} \sim \frac{(\alpha+\epsilon)n(n-1)}{2}$ for some $\epsilon \neq 0$.

We shall show that the first case occurs with probability which is exponentially small in n, whereas the other two occur with probability which is exponentially small in n^2.

We first consider the case where the large deviations in the number of reds is above the mean. Concerning the first way of getting a large deviation, recall that $N_1 \sim \text{Bin}(n, 1/2)$ where N_1 is the number of reds. By the classical theory, (using Theorem 5.1)

$$P\{N_1 \geq an\} = e^{-nl(a)+o(n)} \text{ for } \frac{1}{2} < a < 1$$

for $l(a) = a\log(a) + (1-a)\log(1-a) + \log(2)$; the main point is that this expression is exponentially small in n, rather than in n^2. Now, since the

function $x \to x(1-x)$ is monotone increasing for $x \in [0,1/2]$ and is monotone decreasing for $x \in [1/2,1]$, if we have at least an reds and so at most $(1-a)n$ blues, we have at most $a(1-a)n^2$ potential red-blue edges and so at least $(n^2-n)/2 - a(1-a)n^2$ potential same-same edges; thus, conditional on this large deviation in the number of reds, we get at least

$$((n^2-n)/2 - a(1-a)n^2)p + a(1-a)n^2q) + o(n(n-1)/2) \ (*)$$

edges if $p > q$, and at most $(*)$ edges if $q > p$, with probability tending to 1 as n goes to infinity. (Note that as a ranges over $[0,1]$, the expression $((n^2-n)/2-a(1-a)n^2)p+a(1-a)n^2q)$ takes all values between $n(n-1)p/2$ and $n(n-1)\alpha/2$ by continuity, and no other values; this is why this argument works for all c between p and α and not for any other c).

However, by contrast, if we are in the second situation, and so get an reds but some asymptotic number of edges other than $(*)$, then that requires a large deviation in the number of red-blue edges or the number of same-same edges in addition to the large deviation in the number of reds; and as the number of same-same edges and the number of same-different edges are both of order n^2 in n, the probability of this is asymptotically $e^{-\lambda n^2}$ for some $\lambda > 0$. (This claim is clear if the number of reds and the number of blues are both of order n; even if one of them is so small that it is of order $o(n)$, then the number of vertices of the other colour must be $> (1-\epsilon)n$ for any $1 > \epsilon > 0$, so the large deviation in the number of edges still requires a large deviation in the number of edges between two vertices of the other colour, which will occur with the stated asymptotic probability). Hence the probability of getting asymptotically $(*)$ edges is (by the previous remarks about the monotonicity of g) equal to $e^{-nI(a)+o(n)}$.

The only other case to consider is the third one, that is the probability of a large deviation in the number of edges when there is no large deviation in the number of reds; then there are asymptotically $n/2$ reds and blues, and so there are asymptotically $n^2/4$ red-blue edges and $n^2/4$ same-different edges and so the large deviation in the number of edges requires a large deviation in at least one of the number of same-same or the number of same-different edges, which will happen, again by the classical theory, with probability $e^{-\gamma n^2+o(n^2)}$ for some $\gamma > 0$ and this gives the required result.

Finally, observe that the argument for the case when the large deviation in the number of reds is below the mean is identical. •

We can now clarify the intuitive picture of what is going on here. For c in the range between p and $(p+q)/2$, we can get large deviations more

cheaply than usual just by allowing a large deviation in the number of reds (with probability exponentially small in n) to make a large deviation in the number of edges highly likely; and to get a larger deviation, the cheapest thing to do is just modify the numbers of reds and blues further. However, outwith that range, since we can push the number of reds or blues no further, we have to get the large deviation by the more traditional and harder method of getting a large deviation in the number of edges, which has probability negative exponential in n^2. This also illuminates why, in Lemma 5.5 our estimates relied on the seemingly crude procedure of just considering one term (that corresponding to a monochrome colouring for $p > q$, and that corresponding to as close as possible to equinumerous reds and blues for $q > p$); for the above shows that being in this state is in fact necessary for the exponential in n^2 probabilities to control the large deviation probability.

The above shows that we can have, in a reasonably natural situation, $n(n-1)/2$ trials where the probability of a large deviation is only exponentially small in $\sqrt{n(n-1)/2}$, rather than the usual $n(n-1)/2$. It is natural to ask if we can use this potential for getting large deviations in the number of edges more easily than we would do normally, for c between p and $(p+q)/2$, to prove purely graph theoretic results. This is not clear; it is often stated that we want exponentially small bounds on the probability that a graph does not have some property, but in fact it seems that often at least only exponentially small in some power of n will do.

Recall that classically, one way to prove the upper bound in such large deviation inequalities is by using a martingale inequality.

Definition 5.1 *A sequence of random variables S_n is a* **martingale** *with respect to a sequence X_n of random variables if and only if $\mathbf{E} \mid S_n \mid < \infty$ for all n and for all $k \geq 1$ $\mathbf{E}(S_{k+1} \mid X_1, X_2, ...X_k) = S_k$.*

Recall the following upper bound on the probability of a large deviation;

Theorem 5.11 *Suppose S_n is a martingale with respect to the X_i, and that there exists a sequence of real numbers c_i such that $P\{\mid S_i - S_{i-1} \mid \leq c_i\} = 1$ for all i. Then for $x > 0$ we have*

$$P\{\mid S_n - S_0 \mid \geq x\} \leq 2\exp(-\frac{x^2}{2\sum_{i=1}^{n} c_i^2}).$$

Proof. [GS] Theorem 12.2 (3). •

For example, in the classical model, with $X_{ij} = 1$ if the edge $i - j$ arises, 0 otherwise, and $Y_{ij} = X_{ij} - \alpha$, $\mathbf{E}Y_{ij} = 0$ and the Y_{ij} are independent, so defining $S_n = \sum_{1 \le i < j \le n} Y_{ij}$, we easily see that S_n is a martingale and that $| S_n - S_{n-1} | \le \max\{\alpha, 1 - \alpha\}$. Hence, by Theorem 5.11, we get

$$P\{S_n \ge \frac{an(n-1)}{2}\} \le 2\exp\left(-\frac{a^2n^2(n-1)^2}{4n(n-1)c^2}\right)$$

where $c = \min\{\alpha, 1 - \alpha\}$. However we know that for large enough n, no such inequality can hold in $G_{p,q}$ for a between p and α, as the probability of a large deviation is exponentially small in n rather than n^2. Of course, for the X_{ij} themselves, this is because the next indicator is in general no longer independent of the previous one, which can be shown more elementarily, e.g by Theorem 2.17; however what our argument proves is the slightly stronger fact that there is no martingale Y_n which converges to the number of edges minus the expected number. It seems likely that similar slight imbalances between the numbers of reds and blues will stop us applying martingale techniques in various similar circumstances.

It is clear that we can use the ideas of this section to get random graph models where the probability of a large deviation in the number of edges is negative exponential in arbitrarily small powers of n by having a hierarchy of correlation structure. For example, if $n = m^2$ we might generate m^2 independent random variables, taking values 0 and 1, indexed by ordered pairs (i, j), $0 \le i \le m - 1$ and $1 \le j \le m$, and then say that vertex k is red if both entries of (i, j), where $k = im + j$, are zero or both are one, and is blue otherwise; then a large deviation in the number of the is (say) which are 1, which will happen with probability exponentially small in $m \sim (n(n-1)/2)^{1/4}$ will give a large deviation in the number of reds or blues, and so a large deviation in the number of edges with that probability; and the idea can obviously be iterated to get arbitrarily small powers of n. Note of course that this setup would not have the good independence properties of our models; in particular the colours of the vertices will be correlated.

One thing we have not confirmed yet, strictly speaking, is that the cases which lead to large deviations above are the only cases which lead to large deviations; some more insight into this will follow in the next section, when we consider what happens for general RRC models.

5.3 Large deviations; the general case

We next consider how generally the argument for the large deviations above can be made to work. Crudely speaking the same argument, considering a monochrome colouring to get one bound, and a colouring with as close as possible to the expected number of vertices of each colour to get the bound in the other direction, will still give a rate function, but this function is unfortunately much less explicit than in the previous argument, and we will not be able to take the argument much beyond that point; however some partial insights will be possible.

Theorem 5.12 *In a $\Gamma(n,k,P,\mathbf{s})$ RRC model, considering the probability of a large deviation in the number of edges, the function $\phi(\theta)$ in the Gartner-Ellis theorem is the maximum of the quadratic form*

$$\sum_{i,j=1}^{k} s_i s_j a_{ij} \text{ where } a_{ij} = \log\left(p_{ij}e^{\theta} + 1 - p_{ij}\right)$$

over the simplex

$$\Delta_n = \{(s_1, \ldots s_k) \text{ such that } s_i \geq 0 \text{ for all } i \text{ and } \sum_{i=0}^{k} s_i = 1\}.$$

Proof. Recall that in $\Gamma(n,k,s,P)$, by Theorem 4.11 the moment generating function of $\mathcal{E}_n$ is

$$\sum \binom{n}{j_1, \ldots j_k} \prod_{l=1}^{k} s_l^{j_l} \prod_{1 \leq r < s \leq k} \left(p_{rs}e^{\theta} + 1 - p_{rs}\right)^{j_r j_s} \prod_{r} \left(p_{rr}e^{\theta} + 1 - p_{rr}\right)^{j_r(j_r-1)/2}$$

the sum being over j_i, $1 \leq i \leq k$, such that $j_i \geq 0$ for all i and $j_1 + \ldots + j_k = n$. Let $a_{rs} = \log\left(p_{rs}e^{\theta} + 1 - p_{rs}\right)$, we have

$$\phi_n(\theta) = \frac{2\log(\mathrm{E}(e^{\theta \mathcal{E}_n}))}{n(n-1)}$$

$$= \frac{2\log\left(\sum \binom{n}{j_1, \ldots j_k} \prod s_l^{j_l} \prod \left(p_{rs}e^{\theta} + 1 - p_{rs}\right)^{j_r j_s} \prod \left(p_{rr}e^{\theta} + 1 - p_{rr}\right)^{j_r(j_r-1)/2}\right)}{n(n-1)}$$

the sum being over the same indices as last time, the first product from the left being over l ranging from 1 to k, the second over $1 \leq r < s \leq k$ and the last over $1 \leq r \leq k$. This is

$$\frac{= 2\log(\sum_{j_1+\ldots+j_k=n} \binom{n}{j_1,\ldots j_k} \prod_{l=1}^{k} s_l^{j_l} \prod_{1\leq r<s\leq k} e^{a_{rs}j_r j_s} \prod_{1\leq r\leq k} e^{\frac{a_{rr}j_r(j_r-1)}{2}})}{n(n-1).}$$

Clearly the quadratic form $f(x) = \sum_{i,j=1}^{k} x_i x_j a_{ij}$ attains its maximum on the compact set $\Delta_k = \{(x_1, \ldots x_k) : x_i \geq 0 \forall i, \sum_{i=1}^{k} x_i = 1\}$ at some point $\mathbf{m}$ (which may well be far from unique). We now show that the function ϕ is intimately tied up with such maxima.

Writing $\mathbf{j}$ for $(j_1, \ldots j_k)$, we have that the above is

$$= \frac{2\log(\sum_{j_1+\ldots+j_k=n} \binom{n}{j_1,\ldots j_k} \prod_{l=1}^{k} s_l^{j_l} e^{\frac{f(\mathbf{j})-\sum_{r=1}^{k} j_r a_{rr}}{2}}}{n(n-1)}$$

$$\leq \frac{2\log\left(\sum_{j_1+\ldots+j_k=n} \binom{n}{j_1,\ldots j_k} \prod_{l=1}^{k} s_l^{j_l} e^{\frac{n^2 f(\mathbf{m})-\sum_{r=1}^{k} j_r a_{rr}}{2}}\right)}{n(n-1)}$$

$$= \frac{n^2 f(\mathbf{m})}{n(n-1)} + \frac{2\log \sum_{j_1+\ldots+j_k=n} \binom{n}{j_1,\ldots j_k} e^{j_i\left(\log(s_i)-\frac{a_{ii}}{2}\right)}}{n(n-1)}$$

$$= \frac{n^2 f(\mathbf{m})}{n(n-1)} + \frac{2\log(h^n)}{n(n-1)}$$

where h is the multivariate moment generating function of a multinomial distribution with k equiprobable colours, evaluated at Θ where

$$\theta_i = \log\left(s_i - \frac{a_{ii}}{2}\right);$$

in particular it is independent of n and so the logarithm of h^n is order of magnitude n and so is comfortably killed by the $n(n-1)$ denominator. Thus we have shown that

$$\limsup_{n\to\infty} \phi_n(\theta) \leq f(\mathbf{m}).$$

In the other direction, we can take the term with j_r as close as possible to nm_r subject to the restriction $\sum j_r = n$. Then j_r and nm_r are asymptotically the same; hence, for n sufficiently large,

$$\mathbf{E}(e^{\theta\mathcal{E}_n}) \geq \binom{n}{j_1, \ldots j_k} \prod s_l^{j_l} e^{f(\mathbf{m})n^2(1+o(1))}$$

$$\Rightarrow \liminf_{n\to\infty} \phi_n \geq f(\mathbf{m})$$

since Stirling's formula shows that the logs of the multinomial probabilities are of order of magnitude n at most, so that they are killed by the n^2 term on the bottom. •

We are thus clearly interested in questions about the maxima over Δ_k of $\mathbf{s}^T A\mathbf{s}$. There are two aspects of this; one is trying to find out what this maximum is, and the second is asking which colourings lead to this maximum. We start with the first question. It is not clear how to simplify the expression for ϕ in Theorem 5.12; however one observation can be made.

Theorem 5.13 *Suppose A is a symmetric k by k matrix, so that its eigenvalues are real. Then* $\max_{\mathbf{s}\in\Delta_n} \mathbf{s}^T A\mathbf{s} \leq \mu_1$.

Proof. By the symmetry, A has a basis of eigenvectors $\{\mathbf{e}_i\}$ $1 \leq i \leq k$, which are orthonormal with respect to the usual inner product $(,)$ which satisfy $A\mathbf{e}_i = \mu_i \mathbf{e}_i$ where $\mu_1 \geq \mu_2 \ldots \geq \mu_k$. Let $\mathbf{m}$ be an vector in Δ_k which gives the maximum (which is attained as Δ_k is compact); then, for some choice of λ_i, we have

$$\mathbf{m} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \Rightarrow A\mathbf{m} = \sum_{i=1}^{k} \lambda_i \mu_i \mathbf{e}_i \Rightarrow \mathbf{m}^T A\mathbf{m} = \sum_{i=1}^{k} \lambda_i^2 \mu_i.$$

But also, again by the orthonormality of the eigenvectors, we have that

$$\sum_{i=1}^{k} \lambda_i^2 = (\mathbf{m}, \mathbf{m}) = \sum_{i=1}^{k} m_i^2 \leq 1$$

since $\mathbf{m} \in \Delta_k$. Hence we have

$$\mathbf{m}^T A\mathbf{m} = \sum_{i=1}^{k} \lambda_i^2 \mu_i \leq \sum_{i=1}^{k} \lambda_i^2 \left(\max_j \mu_j\right) \leq \max_j \mu_j \text{ as required.} \bullet$$

In fact we need only consider the submatrix of A corresponding to those rows and columns i for which $m_i \neq 0$. By the interlacing lemma of linear algebra, if A is a symmetric matrix, B is A with the ith row and column omitted, and $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_k$ and $\mu_1 \geq \mu_2 \ldots \geq \mu_{k-1}$ are the eigenvalues of A and B respectively, then $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \ldots \geq \lambda_{k-1} \geq \mu_{k-1} \geq \lambda_k$.

It is natural to ask for a lower bound also, but no good bound of this type is obvious.

We next address the question of when the upper bound is attained. To get $\sum_{i=1}^{k} \lambda_i^2 \mu_i = \mu_1$ when $\sum_{i=1}^{k} \lambda_i^2 \leq 1$ we must have that $\lambda_1 = \pm 1$ with all the other λ_i being zero; as $\lambda_1 \geq 0$, it is 1. Then $\mathbf{e}_1 = (e_1, ...e_k)^T$ must satisfy $\sum_{i=1}^{k} e_i = 1$ as it is in Δ_k, and $\sum_{i=1}^{k} e_i^2 = 1$ as it is an orthonormal vector; thus $\sum_{i=1}^{k} e_i(1 - e_i) = 0$; as $1 \geq e_i \geq 0$ for $\mathbf{e}_1$ in Δ_k, we deduce that one e_i is 1 and the rest are zero.

We now consider questions of which colourings give rise to the maximum. This is the subject matter of much of ESS theory, the basic definitions from which we recall.

Definition 5.2 *Let A be a real k by k matrix. Then an* **evolutionarily stable strategy (ESS)** *of A is an $\mathbf{p} = (p_1, ...p_k) \in \Delta_k$ such that*

$$1. \mathbf{p}^T A \mathbf{p} \geq \mathbf{q}^T A \mathbf{p} \forall q \in \Delta_n.$$

$$2. \textit{If } \mathbf{q} \neq \mathbf{p} \in \Delta_n \textit{ and } \mathbf{p}^T A \mathbf{p} = \mathbf{q}^T A \mathbf{p} \textit{ then } \mathbf{p}^T A \mathbf{q} > \mathbf{q}^T A \mathbf{q}.$$

The curious name arises from game theory in biology; if two populations compete for limited resources, with a finite number of different **pure strategies** each individual can adopt, and if an individual of the first population playing strategy i against an individual from the second playing j receives a payoff a_{ij}, then an ESS $\mathbf{s}$ is a deployment of the population which maximises the overall payoff; we can think of s_k as the probability that an individual should play pure strategy k in any particular conflict. We will consider how many and what kinds of ESSs A can have. Note that there are matrices with no ESSs and others with several ESSs.

Definition 5.3 *The* **support** $R(\mathbf{s})$ *of $\mathbf{s} \in \Delta_k$ is $\{i : s_i \neq 0\}, 1 \leq i \leq k$.*

The **pattern of ESSs** *for A is the set of supports of all the ESSs of A (a subset of the power set of $\{1, 2, ...k\}$). A set of subsets of $\{1, 2, ...k\}$ which is the pattern of ESSs for some matrix A is an* **attainable** *pattern.*

Example 1. If there is some j such that $a_{jj} > a_{ij} \forall i \neq j$, then the j-th pure strategy is obviously an ESS; in fact, it is the only ESS with $s_j > 0$. Such a strategy is said to be **diagonally dominant**.

Example 2. If A is symmetric, an ESS of A is just a maximum of the quadratic form represented by A on the simplex. For example, in genetics, with k alleles $A_1, ...A_k$ and genotype A_iA_j having viability a_{ij}, classical theory

shows that, if the allelic frequency is $\mathbf{s}$ in one generation and $\mathbf{s}'$ in the next, where

$$s_i' = \frac{s_i(A\mathbf{s})_i}{\mathbf{s}^T A\mathbf{s}} \text{ for all } 1 \leq i \leq k$$

the mean fitness $V = \mathbf{s}^T A\mathbf{s}$ is non-decreasing from one generation to the next, and is constant only at an equilibrium point; one proof of this uses Theorem 2.25. Hence for a symmetric matrix, if we start from an $\mathbf{s}$ all of whose components are positive, the process converges to a local maximum, and every such vector is a locally stable equilibrium of the equation above.

Note that thus if $\Gamma(n, k, \mathbf{s}, P)$ is a TID model, $\mathbf{s}' = \mathbf{s}$ as $s_i \neq 0 \Rightarrow (P\mathbf{s})_i = 0$; so $\mathbf{s}$ is an equilibrium point of the system. However it need not be a maximum, as $\mathbf{s} = (1/2, 1/2)^T$ and $G_{p,q}$ with $p > q$ makes clear.

Theorem 5.12 shows that it is of interest to determine what patterns of ESSs a matrix can have. There are some simple combinatorial restrictions;

Lemma 5.14 *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be ESSs of* A, $R(\mathbf{p})$ *be the support of* $\mathbf{p}$ *and* $S(\mathbf{q}) = \{\mathbf{m} \in \Delta_k : \mathbf{m}^T A\mathbf{p} = \mathbf{p}^T A\mathbf{p}\}$ *so that* $R(\mathbf{p}) \subset S(\mathbf{p})$. *Then* $R(\mathbf{q}) \not\subset S(\mathbf{p})$ *and* $R(\mathbf{p}) \not\subset S(\mathbf{q})$. *In particular, for ESSs* $\mathbf{p} \neq \mathbf{q}$ *of* A, $R(\mathbf{q}) \not\subset R(\mathbf{p})$.

Proof. This result is in [BC]. •

Definition 5.4 *A set* $\Im$ *of subsets of* $\{1, 2, ...n\}$. *is a* **Sperner family** *if* $A, B \in \Im$ *and* $A \neq B \Rightarrow A \not\subset B$ *and* $B \not\subset A$.

For example, the previous result makes it clear that an attainable pattern of ESSs is a Sperner family.

Theorem 5.15 *Let* $\Im$ *be a Sperner family of subsets of* $\{1, 2, ...n\}$. *Then*

$$| \Im | \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Proof. This is Sperner's theorem; see [B2] or [A] for much discussion. •

Of course by Stirling's formula the upper bound on the size of $| \Im |$ is about $2^n/\sqrt{2\pi n} = o(2^n)$ as $n \to \infty$. In fact further restriction on attainable patterns exclude certain Sperner families; for example, the triangle exclusion rule [CV] says that if $X \supset \{1, 2, 3\}$ then not all of $X - \{1\}$, $X - \{2\}$ and $X - \{3\}$ can be supports of ESSs. [BCV] gives further information on this, and the maximum number of ESSs there can be in a pattern, including the

fact that it grows exponentially with k at a rate between $30^{1/9} \simeq 1.459...$ and 2 (note these results are for symmetric matrices).

We next check that Theorem 5.12 agrees with Corollary 5.8 and Theorem 5.10 for $G_{p,q}$. Putting $\mathbf{s} = (s, 1-s)$, and maximising $f(s) = \mathbf{s}^T A \mathbf{s}$ by calculus, we find (for $p \neq q$, the only case of interest) $f^{(1)}(s) = 0 \Leftrightarrow s = 1/2$ for $s \in (0,1)$, and that $f^{(2)}(s) = 4((pe^\theta + 1 - p) - (qe^\theta + 1 - q))$ is negative if and only if $(p-q)(e^\theta - 1) < 0$; for $\theta > 0$ this is if and only if $p > q$, and for $\theta < 0$ if and only if $p < q$; thus in these cases the maximum is for $s = 1 - s = 1/2$. In this case we easily check that $\mathbf{s}^T A \mathbf{s} = \alpha e^\theta + 1 - \alpha$. Otherwise the maximum occurs at $s = 0$ or 1; whichever we take we get $\phi = pe^\theta + 1 - p$ as required.

We can approach the result when $p > q$ in another way;

Theorem 5.16 *Suppose* $\exists\, r_0$ *such that* $p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0} = \max\left(p_{rs} e^\theta + 1 - p_{rs}\right)$ *in* $\Gamma(n, k, s, P)$. *Then* $\phi(\theta) = \log\left(p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}\right)$.

Proof. This is very reminiscent of Lemma 5.5. We have

$$\phi_n(\theta) = \log\left(p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}\right)$$

$$+ \frac{2\log(\sum \binom{n}{j_1,\ldots j_k} \prod_{l=1}^k s_l^{j_l} \prod_{1 \leq r < s \leq k} \left(\frac{p_{rs} e^\theta + 1 - p_{rs}}{p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}}\right)^{j_r j_s} \prod \left(\frac{p_{rr} e^\theta + 1 - p_{rr}}{p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}}\right)^{\frac{j_r(j_r-1)}{2}}}{n(n-1)}$$

the sum again being over non-negative integer values of $j_1, \ldots j_k$ such that $\sum_{i=1}^k j_i = n$. By the multinomial theorem and the fact that log is monotonic, this is

$$\leq \log\left(p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}\right).$$

In the other direction, taking only the term with all vertices of colour r_0

$$\phi_{\frac{n(n-1)}{2}}(t) \geq \log\left(p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}\right) + \frac{2\log(\binom{n}{\frac{n}{2}} s_{r_0}^{n_0})}{n(n-1)}$$

which tends to

$$\log\left(p_{r_0 r_0} e^\theta + 1 - p_{r_0 r_0}\right)$$

as required (again exploiting Stirling as in the above proofs to show that the logarithm of the binomial coefficient is only order of magnitude n). •

Yet another proof for $p > q$ uses the fact that the maximum corresponds to an ESS and the fact that $p > q$ implies A is diagonally dominant.

Again note that the argument shows up one way in which large deviations are overwhelmingly likely to arise, i.e through a monochromatic colouring. However it is clear that in general, the possibility of multiple ESSs means that there may be many colourings which give the maximum.

Another question is how the patterns of ESSs change under perturbation of the coefficients of the matrix A (both perturbations from changing θ and changing the entries of P whilst preserving α). Note that the entries of A are a nice smooth function of θ. Broom (personal communication) has initiated the study of how the pattern of ESSs can change as the matrix varies over the space of k by k matrices, but the technical restrictions required for his ideas to work (namely, that the two regions of the space of matrices giving rise to the two different patterns should have a common boundary of codimension 1) are inconvenient for us since we see no reason to suppose that they hold.

It is more or less obvious that here, as in $G_{p,q}$ getting large deviations in the number of vertices of some colour will provide large deviations on the cheap in the number of edges, and thus that the rate function will be uninformative for certain ranges of the parameter values. Indeed in general there will be several such intervals where this problem, and the associated one of the function ϕ not being differentiable arise. This is likely to lead to further complications.

6 Connectedness and connectivity in RRC graphs

6.1 A formula for the probability of connectedness

In this chapter we discuss when random graphs in our models are connected, and if so, what their connectivity is. We first illustrate briefly how one could obtain an exact formula for the probability of connectedness. We then consider the threshold in the classical model, and use knowledge of the classical model to estimate the probability of connectedness in $G_{p,q}$ if one of p and q is small. Then we discuss connectivity.

There are two ways of computing exactly the probability of connectedness in the classical case. One, due to Takacs [Ta] solves the more general problem of finding the distribution, for a random graph $G(n,p)$ with a given set S of vertices, of the number of the n vertices in the component of a vertex in S, by turning the problem into one about queuing theory. This approach seems hard to generalise to our models, since key independence properties are lost. Thus we examine the other approach, an iterative one, mimicking the following theorem of Gilbert, which is easily proven considering the probability that the component containing the vertex 1 has order k.

Theorem 6.1 *Let p be given, and $P_n = P\{G(n,p)$ is connected$\}$. Then $P_n = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} P_k (1-p)^{k(n-k)}$.*

Proof. [B], p177, Exercise 1. •

The result has already been generalised to the case of independent edges arising with possibly different probabilities;

Theorem 6.2 *Let $p_{ij} = P\{$the edge $i-j$ arises$\}$ in a graph on n labelled vertices, and P_n be the probability that such a graph is connected. Then*

$$P_n = 1 - \sum_{Y \subset \{1,2,..n\} \backslash k} P\{G \text{ on } \{1,2...n\} \backslash Y \text{ connected}\} \prod_{i \in \{1,2,...n\} \backslash Y, j \in Y} (1 - p_{ij})$$

where k is any element of $\{1,2...n\}$.

Proof. See [Ke]. •

(If all the p_{ij} are equal to p this formula reduces to Gilbert's). This shows that we can calculate the probability of connectedness in our models by conditioning on the various possible colourings and applying this formula for each case arising. This is a daunting task however, even when we have

noted that only the numbers of vertices of each colour are really important, especially if the number of colours k is large. Thus good asymptotic results are even more necessary here than classically.

6.2 Variability in $P\{G_{p,q}$ connected$\}$.

We first give a simple example to show that the probability of connectedness may vary substantially in our models as compared with the corresponding classical model. The following simple example illustrates this.

Theorem 6.3 *$P\{G_{p,q}$ connected$\}$ takes all values in $[2^{-(n-1)}, 1-2^{-(n-1)}]$ as p and q vary with $p+q=1$.*

Proof. If $q=0$ and $p=1$ then G is connected if and only if all n vertices are the same colour, which happens with probability $2^{-(n-1)}$. If $p=0, q=1$, each vertex is adjacent to every vertex of the opposite colour so the probability of connectedness is the probability that the graph is not monochrome, which is $1-2^{-(n-1)}$. The result follows by the intermediate value theorem, since $P\{G_{p,q}$ is connected$\}$ is a polynomial in p and q so a continuous function. •

This suggests very different behaviour for large p and large q. However, we now prove a result limiting the extent of asymmetry between p and q.

Theorem 6.4 *Let $A \subset V(G)$, $\emptyset \neq A \neq V(G)$. Then, if $E(A,B)$ is the set of edges between the two sets of vertices A and B*

$$P\{E(A,A^c)=\emptyset \text{ in } G_{p,q}\} \geq P\{E(A,A^c)=\emptyset \text{ in } G_\alpha\}.$$

and the expression on the left-hand side is symmetric between p and q. There is equality if and only if $p=q$ or $|A|$ or $|A^c|$ is 1.

Proof. If $|A|=i$, then $P\{E(A,A^c)=\emptyset \text{ in } G_{p,q}\}$

$$= \sum_{j=0}^{i} P\{E(A,A^c)=\emptyset \mid N_1(A)=j\} P\{N_1(A)=j\}$$

$$= \sum_{j=0}^{i} \left(\frac{1}{2}\left((1-p)^j(1-q)^{i-j}+(1-q)^j(1-p)^{i-j}\right)\right)^{n-i} \frac{\binom{i}{j}}{2^i}$$

since if there are j red vertices in A, then for any element not in A, if it is red the probability that it is joined to no element of A is $(1-p)^j(1-q)^{i-j}$

and if it is blue the probability is $(1-q)^j(1-p)^{i-j}$. This expression is symmetric between p and q as required. To prove the inequality, recall that since $f(x) = x^{n-i}$ is a strictly convex continuous function of x, for $a_i \geq 0$ with $\sum_{i=1}^{k} a_i = 1$, we have $f(\sum_{i=1}^{n} a_i x_i) \leq \sum_{i=1} a_i f(x_i)$ with equality if and only if f is linear or the x_j take only one value. The result follows noting that

$$\left(\sum_{j=0}^{i} \frac{\binom{i}{j}}{2^i} \left(\frac{(1-p)^j(1-q)^{i-j} + (1-q)^j(1-p)^{i-j}}{2}\right)\right)^{n-i}$$

$$= \left(\frac{1-p}{2} + \frac{1-q}{2}\right)^{i(n-i)} = (1-\alpha)^{i(n-i)} \text{ as required.} \bullet$$

Note that the symmetry depends on the fact that we only consider two sets; with three non-empty disjoint sets A, B, C involved, it is easy to show that $P\{E(A,B) = E(B,C) = E(A,C) = \emptyset\}$ is asymmetric between p and q. Also it is clear that asymmetry between p and q enters into the formula only through the probability that $G_{p,q}$ has ≥ 3 components; for, by the inclusion-exclusion formula, we have

$$P\{G \text{ connected}\} = 1 - \sum_{A \subset V(G), V(G) \neq A \neq \emptyset} P\{E(A, A^c) = \emptyset\}$$

$$+ \sum_{A \neq B \subset V(G)\, V(G) \neq A, B \neq \emptyset} P\{E(A, A^c) = E(B, B^c) = \emptyset\} - \ldots$$

The first non-constant term is symmetric by Theorem 6.3 so $P\{G_{p,q}$ is connected$\}$ is asymmetric between p and q only if some higher term in the formula is non-zero. This means we must have at least three components since if

$$P\{E(A, A^c) = E(B, B^c) = \emptyset\} > 0 \text{ with } A \neq B \subset V(G),\ V(G) \neq A, B \neq \emptyset$$

at least three of $A \cap B$, $A \cap B^c$, $A^c \cap B$ and $A^c \cap B^c$ must be non-trivial.

6.3 Asymptotics for $P\{G_{p,q}\}$ connected.

We now recall the following result, which describes the asymptotics of the probability of connectedness in the classical case.

Theorem 6.5 *Suppose* $\alpha = (\log(n) + c + o(1))/n$. *Then*

$$\lim_{n \to \infty} P\{G_\alpha \textit{ is connected}\} = e^{-e^{-c}}.$$

Proof. See [B], Theorem VII.3 •.

Thus $\log(n)/n$ is the threshold for the probability of connectedness. Some remarks on the proof will enable us to introduce ideas which will be useful in the sequel. The main point is that, for α sufficiently large, a.e. graph consists of a so-called giant component and some isolated vertices. Thus the probability of connectedness is, in the limit, the probability that no vertex is of degree 0. As we have already seen, the degrees are dependent, even in the classical model; however, for α as in the theorem, they behave asymptotically as if they are independent in various ways; in particular, the standard extreme value analysis for the minimum of n independent binomials is applicable, which is why we get the familiar distribution on the right hand side.

Note that some insight into the analogous problem for random graphs where edges arise independently but with possibly differing probabilities can be had from the following result;

Theorem 6.6 *Let p_{ij} be the probability of the edge $i-j$ in a random graph on $\{1,2,...n\}$, the various edges being independent. Let $q_{ij}=1-p_{ij}$ and $Q_{ir}=\max_{j_1<...j_{n-r}} q_{ij_1}q_{ij_2}...q_{ij_{n-r}}$ and $\lambda=\sum_{i=1}^{n}Q_{i0}$ so that λ is the expected number of isolated vertices. If*

$$\lim_{n\to\infty}\max_{1\le i\le n}Q_{i0}=0 \text{ and } \lim_{n\to\infty}\lambda=\lambda_0 \text{ and } \lim_{n\to\infty}\sum_{t=1}^{\lfloor n/2\rfloor}\frac{\left(\sum_{i=1}^{n}Q_{ir}\right)^r}{r!}-\left(e^{\lambda}-1\right)=0$$

then a.e. such graph consists of a giant component and isolated vertices, whose number converges in distribution to a $Po(\lambda_0)$ random variable.

Proof. [Ko] •

It is not easy to check the conditions of the theorem, especially the last one, though that condition holds if every $p_{ij}=(\log(n)+O(1))/n$ [B, p177]. A problem with attempting to apply Kovalenko's result in our situation is that we get different models according to the random number of vertices of each colour and it is not clear how to put the results together.

Note that the expected number of isolated vertices in $G_{p,q}$ is the same as the number in G_α, as a vertex is isolated if and only if the tree comprising all edges out of that vertex arises in the complement of the graph, so the probability of the event is the probability of that tree in $G_{1-p,1-q}$ which, as $G_{p,q}$ is TID is the same as classically; now use linearity of expectation. However other moments will differ from classically.

To clarify the notion of giant component, we quote the following theorem.

Theorem 6.7 *For a.e. graph with at least* $\lfloor (n/2) + 2\log(n)\, n^{2/3} \rfloor$ *edges, there is a unique* **giant** *component with at least* $n^{2/3}$ *vertices, all other components having at most* $n^{2/3}/2$ *vertices.*

Proof. [B] Theorem VI.1 •

In our models, it is quite possible that we do not have a unique giant component; for example, with k colours, if the p_{ii} $(1 \leq i \leq k)$ are large enough to ensure there is a giant component amongst the vertices of colour i, there can be several large components if all p_{ij}, $i \neq j$ are small enough.

We now move towards a generalisation of Theorem 6.5 to $G_{p,q,r}$ when $\alpha = (\log(n) + c)/n$ and q is sufficiently small. The proof will be heavily dependent on the above results on the structure of the random graph in the classical model. We first note that in our circumstances we can refine the result on existence of the giant component.

Theorem 6.8 *If* $\alpha = (\log(n) + c + o(1))/n$, *then a.e.* G_α *consists of a giant component and isolated vertices, whose number converges in distribution as* n *goes to infinity to a* $Po(e^{-c})$ *random variable.*

Proof. The first statement is demonstrated in the course of the second proof of Theorem VII.3 in [B] (page 151). The second sentence is Theorem V.3 in [B] (page 94) in the special case $k = 1$. •

For the next lemma, it will be helpful to recall the generalisation of the product formula for the exponential in Lemma 4.19.

Lemma 6.9 *Let* $c_n \geq 0$ *be a sequence tending to* c. *Then*

$$\lim_{n\to\infty} \sum_{i=0}^{n} \frac{(1 - c_n/n^2)^{i(n-i)} \binom{n}{i}}{2^n} = e^{\frac{-c}{4}}.$$

Proof. We first note that, as $i(n-i) \leq \lfloor \frac{n^2}{4} \rfloor$ for $1 \leq i \leq n$, the quantity to be evaluated is

$$\geq \lim_{n\to\infty} \left(1 - c_n/n^2\right)^{\lfloor \frac{n^2}{4} \rfloor} = e^{-c/4} \text{ by Lemma 4.19.}$$

In the other direction, since for any $1/2 > \epsilon > 0$, the number of reds i satisfies $\lim_{n\to\infty} P\{i(n-i) \geq \left(n^2 (1/2 - \epsilon)^2\right)\} = 1$ and hence

$$\lim_{n\to\infty} \sum_{i=1}^{n-1} \frac{(1 - c_n/n^2)^{i(n-i)} \binom{n}{i}}{2^n}$$

$$\leq \lim_{n\to\infty} \sum_{i=1}^{n-1} \frac{(1-c_n/n^2)^{n^2(1/2-\epsilon)^2}\binom{n}{i}}{2^n} = e^{-c(1/2-\epsilon)^2}$$

by Lemma 4.19. Letting ϵ go to zero now gives the result. •

We can now go to our main result.

Theorem 6.10 *Suppose we are in $G_{p,q,r}$ with*

$$p = \frac{2\log(n) + c_1 + o(1)}{n} \text{ and } r = \frac{2\log(n) + c_2 + o(1)}{n} \text{ and } q = \frac{c_3}{n^2}$$

where the c_i are constants. Then

$$\lim_{n\to\infty} P\{G_{p,q,r} \text{ is connected}\} = e^{-1/2\left(e^{-c_1/2}+e^{-c_2/2}\right)}\left(1-e^{-c_3/4}\right).$$

Proof. We first note that with probability tending to 1, the numbers of reds and blues are both $n/2 + o(n^{1/2+\epsilon})$ for any $\epsilon > 0$. Hence by Theorem 6.8, again with probability tending to 1, the reds consist of a giant component and isolated vertices, and the same is true of the blues. Because of this, the event that the graph is connected is (up to an error probability which tends to zero) the disjoint union of the following two events;

1. The number of components in the reds is one and the number of components in the blues is one, and there is an link between these.

2. There is at least one vertex which is isolated in the reds or the blues but is joined to the other vertices by adding red-blue edges.

We first claim the probability of the second event tends to zero as n goes to infinity. Indeed for any particular red vertex, the probability that it is not joined to any of the blue vertices is

$$\left(1-\frac{c_3}{n^2}\right)^{n(1/2+o(1))}$$

which clearly goes to 1 as n goes to infinity. Thus the overall probability of the second event is a sum, over the Poissonly distributed numbers of vertices isolated in their own colour, of events with probability tending to 0 with n, and so itself tends to zero.

We may thus concentrate our attention on the first event, A say. It itself is the intersection of three events; the reds being connected, the blues being

connected, and there being a red-blue edge. We consider the limit, as n goes to infinity, of the probabilities of each of these three events in turn; as we will see, given that there are $n/2 + o(n^{1/2+\epsilon})$ reds, the answers do not depend on the exact number of reds, and so, since the events are independent conditional on the number of reds, it is enough to work out the individual probabilities and multiply them together. We start with the probability that the reds are connected. We have

$$p = \frac{2\log(n) + c_1 + o(1)}{n} = \frac{\log(n/2) + \log(2) + c_1/2 + o(1)}{n/2}$$

$$= \frac{\log(n/2 + o(n^{1/2+\epsilon})) + \log((n/2)/(n/2 + o(n^{1/2+\epsilon}))) + \log(2) + c_1/2 + o(1)}{n/2}$$

$$= \frac{\log(n/2 + o(n^{1/2+\epsilon})) + \log(2) + c_1/2 + o(1)}{n/2}$$

$$= \frac{\log(n/2 + o(n^{1/2+\epsilon})) + \log(2) + c_1/2 + o(1)}{n/2 + o(n^{1/2+\epsilon})}$$

where we use the fact that $\lim_{n\to\infty} \log(n)/n^\kappa = 0$ for any $\kappa > 0$ to mop up the extra terms caused by the above changes into the $o(1)$ term. Hence by Theorem 6.5

$$\lim_{n\to\infty} P\{\text{the reds are connected}\} = e^{-e^{-(\log(2)+c_1/2)}} = e^{-1/2e^{-c_1/2}}.$$

The proof that the

$$\lim_{n\to\infty} P\{\text{the blues are connected}\} = e^{-1/2e^{-c_2/2}}$$

is identical. Hence it remains to get a good estimate of the probability that there is a red-blue edge, conditional on the numbers of blues and reds being as above. But we have just seen that this varies from the exact probability that there is a red-blue edge by at most an error term tending to zero as n goes to infinity. Hence we may approximate it by the exact answer, which is

$$P\{\text{there is a red-blue edge}\} = 1 - P\{\text{no red-blue edge}\}$$

$$= 1 - \sum_{i=0}^{n} \frac{(1 - c_3/n^2)^{i(n-i)}\binom{n}{i}}{2^n}$$

and hence we can apply Lemma 6.9 to see that

$$\lim_{n\to\infty} P\{\text{a red-blue edge}\} = 1 - e^{-c_3/4}.$$

Consequently, putting all the strands together,

$$\lim_{n\to\infty} P\{G_{p,q,r} \text{ is connected}\} = e^{-1/2e^{-c_1/2}} e^{-1/2e^{-c_2/2}} \left(1 - e^{-c_3/4}\right)$$

$$= e^{-1/2\left(e^{-c_1/2}+e^{-c_2/2}\right)} \left(1 - e^{-c_3/4}\right). \bullet$$

In fact we have shown rather more. Since the event described as 2 in the proof has probability tending to zero, with probability tending to one we have that the number of isolated vertices is the sum of the number of vertices isolated in the reds and the number of vertices isolated in the blues. As

$$p = \frac{\log(n/2) + \log(2) + c_1 + o(1)}{n/2} \text{ and } r = \frac{\log(n/2) + \log(2) + c_1 + o(1)}{n/2}$$

these are asymptotically Poisson with parameters $e^{-c_1}/2$ and $e^{-c_2}/2$ respectively, by Theorem 6.8, and the numbers of isolated vertices in the reds and blues are asymptotically independent, we have that the total number of isolated vertices is asymptotically Poisson with parameter $(e^{-c_1} + e^{-c_2})/2$.

Note the correlation structure does show up, for the corresponding classical probability is $e^{-e^{-(c_1+c_2)/4}}$ since, as $q = c_3/n^2 = o(1)$,

$$\alpha = \frac{p + 2q + r}{4} = \frac{\log(n) + (c_1 + c_2)/4 + o(1)}{n}.$$

The argument fails in its present form if the colours are not equiprobable; indeed if $s < 1/2$, we have

$$p = \frac{(2\log(n) + c + o(1))}{n} = \frac{2s\log(sn) + d + o(1)}{sn}$$

where d is a constant whose exact value is immaterial; thus p is less than the critical probability $\log(sn)/sn$ by about $(1 - 2s)\log(sn)/n$, which is more than any particular c/n; hence, again by Theorem 6.5, the probability that the reds are connected is in the limit zero, and the probability that all the red vertices are joined to a blue one (the blues will, by contrast, be connected with probability tending to 1) will also tend to zero, and so the probability

of connectedness will tend to zero. However we can remedy this by breaking down α between p and r in a more appropriate way. Clearly we want, if there are about sn reds and $(1-s)n$ blues, to have

$$p = \frac{\log(sn) + c_1 + o(1)}{sn} \text{ and } r = \frac{\log((1-s)n) + c_2 + o(1)}{(1-s)n}$$

for suitable choices of c_1 and c_2 to get the above argument to work. However then, recalling that $q = o(1/n)$, we have

$$\alpha = s^2p + 2s(1-s)q + (1-s)^2r$$

$$= \frac{s\log(sn) + sc_1 + (1-s)\log((1-s)n) + (1-s)c_2 + o(1)}{n}$$

so we must have $sc_1 + s\log(s) + (1-s)\log(1-s) + (1-s)c_2 = c$. Then an analogous argument to that above will give us the non-degenerate limit

$$\lim_{n\to\infty} P\{{}_sG_{p,q,r} \text{ is connected}\}.$$

It is natural to ask how far these techniques may be used to investigate the probability of connectedness for other values of the probabilities. We consider the case where the classical graph is **sparse**, that is we have $\alpha = c/n$ where c is a constant. The following result of Erdos and Renyi shows that the behaviour of such graphs with respect to their number and order of components is critically dependent on the size of c.

Theorem 6.11 *If $c > 1$, in a.e. $G(n, c/n)$ the largest component has order $N_{n,c}$ such that $\mid N_{n,c} - (1-t(c))n \mid \leq \omega(n)\sqrt{n}$ where*

$$t(c) = \frac{1}{c}\sum_{k=1}^{\infty} \frac{k^{k-1}(ce^{-c})^k}{k!}$$

so that $s(c) = ct(c)$ is the unique root of $se^{-s} = ce^{-c}$ in $(0,1]$; and the other components have orders X such that $(\mid X - (\log(n) - 5\log\log(n)/2) \mid)/\tau \leq \omega(n)$ for any $\omega(n)$ tending to infinity with n, where $\tau = c - 1 - log(c)$. However, if $c < 1$ all components are small, having orders which are $O(\log(n))$.

Proof. The first sentence is [B] Theorem VI.11; the rest is implied by [B] Chapter V and stated explicitly in the introduction to chapter VI. •

This so-called **double jump** phenomenon was one of the earliest examples in random graph theory of a radical change in the structure of the random graph arising from a slight change in the parameter. Much more detailed results are now known about how the transition occurs, see e.g [JKLP].

We now suppose we are in $G_{p,q}$ with $p+q = 2a/n$ and $p = c/n$, $q = d/n$, or perhaps $p = 2a/n - f/n^2$, $q = e/n^2$; what is the behaviour here? Of course if $a < 1/2$ then both c and d are less than 1 so we will only get small components, but (for example) if $c > 1$ the red and blue sets will a.s. have large components and there is interest in firstly whether the two giant components will join up and if so how much else will join up with them.

The first question is approachable. Indeed a.e such graph has a giant red component C_1 with about $\omega(c)n$ vertices, and a similar component C_2 of blues. Thus the probability that there is no edge between them will be, by an argument based again on Lemma 6.9, $(1-q)^{\omega(c)^2(n^2+o(n^2))}$ which by the product formula for the exponential will go to 0 as n goes to infinity if $q = c/n$, but will go to $e^{-f\omega(c)^2}$ if $q \sim f/n^2$.

It seems harder to understand the probability of full connectedness here. Of course it is obvious that this probability will tend to zero, from the nature of the threshold probability for connectedness, but one might hope to understand the rate at which it goes to zero, show that this is the same rate at which the the probability of no isolated vertex goes to zero, and deduce something about the asymptotic form of the probability of connectedness. However O'Connell [O] has recently shown that, if $\alpha = a/n$, the properties of being connected and of having no isolated vertices no longer have the same asymptotic probability in the classical model; more precisely, he shows that, for any $c > 0$, defining $a > 0$ uniquely by $1 - e^{-a} = c/a$,

$$m(c) = \lim_{n\to\infty} \frac{\log(P\{G(n,c/n) \text{ is connected}\})}{n} = \log(1-e^{-c}),$$

$$g(c) = \lim_{n\to\infty} \frac{\log(P\{G(n,c/n) \text{ has no isolated vertex}\})}{n} = \log(\frac{c}{a}) - \frac{(c-a)^2}{2c}$$

and one can show that $g(c) > m(c) \forall c > 0$.

6.4 The limiting probability of connectedness; small p

The other extreme is when $p = 0$, $q = 2(\log(n) + x + o(1))/n$, when with probability tending to one the graph is bipartite, with classes the red vertices

and the blue vertices. The probability that a random bipartite graph is connected seems to have first been addressed by Palasti [Pa]; further results were proved by Klee, Larman and Wright [KLW], and later by Saltyakov [Sa]. The techniques used in all these papers are similar; in particular, a key step in each proof is to prove an analogue of the result of Erdos and Renyi that connectedness is essentially the same as having no isolated vertex. We now give the result of [KLW] and show how it solves our problem.

Theorem 6.12 *Suppose we have a random graph with vertex classes whose orders are $n_1 \leq n_2$ respectively, where $n = n_1 + n_2$. Let g be fixed and γ be bounded, and suppose there are $\mathcal{E} = n_2(n_1 - (n_1 - \gamma)e^{-(\log(n_2))/n_1})$ edges, all such graphs being equally probable. Then, if $\beta = (n_2 - n_1)\log(n)/n \to b$ and $\gamma \to g$ as $n \to \infty$, we have that the probability that the graph is connected tends to $e^{-e^{-g}(1+e^{-b})}$.*

Proof. [KLW] (some notation has been changed to avoid confusion). •

Theorem 6.13 *If $p = 0$, $q = 2(\log(n)+x)/n$, $\lim_{n\to\infty} P\{G_{p,q}$ is connected$\} = e^{-e^{-x}}$.*

Proof. First note that without loss of generality the blues are at least as numerous as the reds; then there are $N_1 = n/2 - o(n^{1/2+\epsilon})$ for any $\epsilon > 0$. Thus $\beta = o(n^{\epsilon-1/2}\log(n))$ so $\beta \to 0$ with n.

Next note that the number of edges will, with probability tending to 1 as $n \to \infty$, be

$$\mathcal{E} = n_1 n_2 \frac{2(\log(n) + x)}{n}(1 + o(n^{\epsilon-1})).$$

Indeed there will be $n_1 n_2$ independent trials, and with probability tending to 1 $n_1 \leq n/2 + \sqrt{n}\omega(n)$ and $n_2 \geq n/2 - \sqrt{n}\omega(n)$ for any $\omega(n)$ tending to infinity with n, we see that the standard deviation of the number of edges produced is a constant times n minus smaller order stuff, which gives the claim. This expression is in turn equal to

$$n_1 n_2 \frac{(\log(n/2) + \log(2) + x)}{n/2}(1 + o(n^{\epsilon-1}))$$

$$= n_1 n_2 \frac{\log(n_2) + \log(2) + x}{n_1}(1+o(n^{\epsilon-1/2})) = n_2(\log(n_2)+(\log(2)+x))+o(n_2)$$

replacing $n/2$ by n_2 and n_1 respectively, and noting for example that the series for $\log(1+x)$ implies $\log(n/2 + c\sqrt{n}) = \log(n/2) + O(n^{-1/2})$.

By Theorem 6.12 it will suffice to show that this is equal to

$$n_2(n_1 - (n_1 - \gamma)e^{-(\log(n_2))/n_1}) \, (*)$$

where $\gamma \to x + \log(2)$. To see this, note that $(*)$ is

$$= n_2 n_1 (1 - e^{-\log(n_2)/n_1}) + n_2 \gamma e^{-\log(n_2)/n_1}$$

$$= n_2 \log(n_2) + O((\log(n_2))^2) + n_2\gamma + O(\log(n_2)) = n_2 \log(n_2) + n_2\gamma + o(n_2)$$

using the Taylor expansion of the exponential; and the last line is the required expression. •

In particular, in the limit the probability of connectedness is the same as in the classical case. This in turn begs the question of how the limiting probability of connectedness varies as q varies in the range from α to 2α; for example, is it constant? The question does not seem easy to approach using the kind of technology we have developed, since it is possible for the whole graph to be connected without the reds being connected, and possible for the whole graph to be connected without the bipartite graph on the reds and blues being connected.

6.5 Connectivity properties of RRC graphs

In the previous sections, we have seen that we have to work quite hard to get the correlation structure showing up in the limit in our models for connectedness. In this section we deal with connectivity, that is the strength of connectedness for a connected graph, where as we shall see rather more satisfactory results seem to be possible.

There are various measures of connectivity. They include the following;

Definition 6.1 *The* **vertex-connectivity** $\kappa(G)$ *is the minimum order of a set of vertices whose removal renders G disconnected.*

The **edge-connectivity** $\lambda(G)$ *is the minimum order of a set of edges whose removal renders G disconnected.*

(Fiedler's) algebraic connectivity $\mu(G)$ *is the second smallest eigenvalue of the Laplacian matrix $\nabla(G)$, where we recall that $\nabla(G) = D - A$ where D is the diagonal matrix with the degree of vertex i in the ith position and A is the adjacency matrix of G, defined by $a_{ij} = 1$ if the vertices i and j are adjacent in G and $a_{ij} = 0$ otherwise.*

Another commonly used measure of connectivity is the minimum degree $\delta(G)$. Closely tied up with connectivity is the diameter of the graph;

Definition 6.2 *The* **diameter** $d(G)$ *of a graph* G *is the maximum, over all pairs* $x \neq y \in V(G)$, *of the lengths of the shortest path in* G *from* x *to* y *(we put* $d(G) = \infty$ *if* G *is not connected).*

It is not hard to see that $\delta \geq \lambda \geq \kappa \geq \mu$. Given $\delta \geq \lambda \geq \kappa$ positive integers, it is possible to exhibit a graph with minimal degree δ, edge-connectivity λ and vertex-connectivity κ, but this behaviour is atypical in the classical model, as the next result makes clear.

Theorem 6.14 *Let a* G_α *evolve, by adding one edge at each time 1,2,..., and let* $k = k(n)$ *be any function from* $\{1, 2, ..n\}$ *to* $\{1, 2, ...(n-1)\}$. *Then in a.e.* G_α *the vertex-connectivity (which is a non-decreasing function of time) first becomes* k *at the same time as the minimum degree first becomes* k. *In particular, for a.e.* G_α, $\delta = \kappa$.

Proof. [B], Theorem VII.4 •

(We will see in Theorem 6.15 below that in G_α with α constant, μ is classically of the same order of magnitude as δ and κ). However the analogue of Theorem 6.14 cannot be true in general in $G_{p,q}$; for if q is very small (order of magnitude $1/n^2$) there will tend to be at least two monochromatic components which may however (e.g if p is constant) have large minimum degrees. Similarily, the fact that in the classical model a.e. graph has a Hamiltonian cycle as soon as it has minimum degree at least two ([B], Theorem VIII.9), cannot be true in such a $G_{p,q}$ model; for although the minimum degree will be (as we have seen in Chapter 4) at least as large as before, the fact that there is no connection between the reds and the blues means there can be no Hamiltonian cycle. This at least suggests that the measures of connectivity are more spaced out in our models; we explore this notion.

The only previous result relating connectivity in our situation is the following, for which a preliminary definition is required.

Definition 6.3 *A sequence* X_n *of random variables is said to be* $o(n^\gamma)$ **in probability** *if* $\forall \delta > 0$ *and* $K > 0$ $\exists n_0$ *such that*

$$n > n_0 \Rightarrow P\{|\frac{X_n}{n^\gamma}| > K\} < \delta.$$

Theorem 6.15 *Let α, p, q and r be constants. Then given $\epsilon > 0$, we have*

$$\mu = \alpha n + o(n^{1/2+\epsilon}) \text{ in probability in } G_\alpha.$$

If we have i reds and $n-i$ blues, with red-red edges arising with probability p, blue-blue edges with probability r and red-blue edges with probability q, then

$$\mu = \theta n + o(n^{1/2+\epsilon}) \text{ in probability, where}$$

$$\theta = \min\{q, \frac{pi + q(n-i)}{n}, \frac{qi + r(n-i)}{n}\}.$$

In particular, if $q < min(r,p)$ $\mu = qn + o(n^{1/2+\epsilon})$.

Proof. [Ju] •

We now use this result to obtain a result for our model.

Theorem 6.16 *In ${}_sG_{p,q,r}$, $\forall \epsilon > 0$,*

$\mu = \theta n + o(n^{1/2+\epsilon})$ *for a.e. graph, where* $\theta = \min\{q, q(1-s)+ps, qs+r(1-s)\}$.

Proof. Given $\epsilon > 0$ a.e. ${}_sG_{p,q,r}$ has between $sn - n^{1/2+\epsilon/2}$ and $sn + n^{1/2+\epsilon}$ red vertices. Hence

$$\frac{pi + q(n-i)}{n} = q(1-s) + o(n^{-1/2+\epsilon}) \text{ and}$$

$$\frac{qi + r(n-i)}{n} = qs + r(1-s) + o(n^{-1/2+\epsilon}).$$

Hence, for a.e. ${}_sG_{p,q,r}$,

$$\theta = \min\{q, \frac{pi + q(n-i)}{n}, \frac{qi + p(n-i)}{n}\}$$

$$= \min\{q, q(1-s) + ps, qs + r(1-s)\} + o(n^{-1/2+\epsilon}).$$

Since $o(n^{-1/2+\epsilon/2})n$ is $o(n^{1/2+\epsilon})$ the result follows. •

Hence, for example, in $G_{p,q}$ if $p > q$, $\theta = q$ but if $q > p$, $\theta = (p+q)/2$; again there is different behaviour for $p > q$ and $q > p$. It is easy to see, simulating graphs in $G_{p,q}$ on 500 vertices, that the observed behaviour of μ is close to that described by the above theory; in fact, this yields the further insight that the connectivity seems to continue to be an increasing function

of q (though less rapidly) for $q > p$; whereas for $q < p$ it increases linearly with q, here it only increases with the square root of q roughly; this will be caused by the slight imbalance in the numbers of reds and blues (which is of course of size order $\sqrt{n}$); causing one of $(pi+q(n-i))/n$ and $(qi+r(n-i))/n$ to be slightly smaller than the other.

Juhasz remarks that his article was motivated by the idea that μ manifests the extent to which the graph is breakable into two blocks, so that a similar result holds in our models is unsurprising.

It seems likely that there is a generalisation to several colours, with the weakest link again controlling the size of μ, but we have not formalised this.

What can we say about the other two measures? We start with λ, using the following lemma.

Lemma 6.17 *If a graph has $\lambda < \delta$ then its diameter is ≥ 3.*

Proof. [Pl] •

This suggests the diameter may give us a grip on when $\lambda \neq \delta$. For this, we shall require information on the diameter.

Theorem 6.18 *Suppose $c > 0$ is real, that $d > 1$ is an integer, and that*

$$\alpha^d n^{d-1} = \log\left(\frac{n^2}{c}\right) \text{ and } \lim_{n\to\infty}\left(\frac{\alpha n}{(\log(n))^3}\right) = \infty.$$

Then

$$\lim_{n\to\infty} P\{d(G_\alpha) = d\} = e^{-c/2} \text{ and } \lim_{n\to\infty} P\{d(G_\alpha) = d+1\} = 1 - e^{-c/2}.$$

In particular, if $\lim_{n\to\infty} p^2 n = \infty$ and $\lim_{n\to\infty} pn - 2\log(n) = \infty$ a.e. $G(n,\alpha)$ has diameter 2.

Proof. [B], Theorem X.10 and Corollary X.11 •

This allows us to deduce something about the behaviour of λ;

Theorem 6.19 *A.e. $G_{p,q}$ with p and q non-zero and constant has $\lambda = \delta$.*

Proof. By Theorem 6.18 a.e. G_α with α constant and non-zero has diameter 2. Clearly the diameter of our $G_{p,q}$ is bounded above by that of $G_{\min\{p,q\}}$ and below by that of $G_{\max\{p,q\}}$; since both these are 2 for a.e. graph, a.e. $G_{p,q}$ has diameter 2. Hence, by Lemma 6.17, a.e. such graph has $\lambda = \delta$. •

Hence, using Corollary 4.22 and the fact that the minimum degree of a.e $G_{p,q}$ is $n-1-\Delta$ where Δ is the maximum degree, λ will be larger than classically for a.e graph. (Of course when $q=0$ a.e. $G_{p,q}$ is disconnected and so has $\lambda=0$ but will still have large δ. Also note the result is for p and q constant; if $q=c/n^2$ with c constant, clearly λ will be close to $c/2$ but δ will still be large).

κ seems to be rather harder to say much about; however it seems highly likely that it will like δ and λ be of about the same order of magnitude as classically, since in a.e. graph every vertex will have about $\alpha n + o(n)$ neighbours.

We close this section with some more remarks about the diameter. We saw in Theorem 6.18 that the case when $\alpha = c/\sqrt{n}$ is where the result on where the diameter being 2 breaks down, so let us analyse what happens then somewhat more closely. Suppose

$$\alpha = \frac{a}{\sqrt{n-2}} \text{ and } p = \frac{a+b}{\sqrt{n-2}} \text{ and } q = \frac{a-b}{\sqrt{n-2}}$$

where of course $-a < b < a$. Let $i \neq j$ be two vertices, and consider

$$H_n(p,q) = P\{d(i,j) > 2 \text{ in } G_{p,q}\}.$$

Now conditional on whether or not the vertices i and j are the same colour, the event $d(i,j) > 2$ is the two independent events that i and j are not adjacent, and that for all other vertices $k \in \{1,2,..n\} - \{i,j\}$ the path $i-k-j$ is not present; thus

$$H_n(p,q) = \frac{1}{2}(1-p)\left(1-\frac{p^2+q^2}{2}\right)^{n-2} + \frac{1}{2}(1-q)(1-pq)^{n-2}$$

so that

$$\lim_{n\to\infty} H_n(p,q) = \frac{e^{-(a^2+b^2)} + e^{-(a^2-b^2)}}{2} = e^{-a^2}\cosh\left(b^2\right)$$

which is symmetric in p and q (unlike H_n), is uniquely minimised by $b=0$ for fixed a, and we can get H_n to be arbitrarily close to $1/2$ by taking b as close to a as we like. (Note that a.e. graph with these probabilities is connected, by Theorem 6.5). Thus the behaviour of the diameter at this critical probability distinguishes between G_α and $G_{p,q}$. Again the symmetry is misleading for the more general question of the probability that two vertices are at distance at least k from each other.

6.6 Isoperimetric inequalities and the concentration of measure.

A consequence of the fact that μ is small in $G_{p,q}$ for small values of q is that, crudely speaking, the graph G will have less good expanding properties. To formalise this, we have the following definition;

Definition 6.4 *A graph G on n vertices is said to be an* (n, c)**-magnifier** *for some $c > 0$ if, letting $N(X)$ denote the set of vertices adjacent to the vertices in X,*

$$\forall X \subset V(G) \text{ such that } \mid X \mid \leq \frac{n}{2} \text{ we have } \mid N(X) - X \mid \geq c \mid X \mid .$$

(An inequality of the type mentioned in the definition, or more generally one relating the number of points at distance at most k from sets A to the order of A is called an **isoperimetric inequality**). This notion is then used in

Theorem 6.20 *If G is an (n, c)-magnifier, then*

$$\mu(G) \geq \frac{c^2}{4 + 2c^2}.$$

Proof. [Al]. •

Sharper results are possible for special classes of graphs, e.g regular graphs.

Theorem 6.16 above on μ in our models thus implies that the graph will (for $p > q$) be a less good expander than classically. Our results also imply that certain functions on the graph will be less tightly concentrated than classically. This we formalise through the following definitions and theorem. It will be helpful to turn G into a probability space by putting the uniform distribution on its vertex set; then a function on G can also be thought of as a random variable.

Definition 6.5 *1. A function $f : V(G) \to \mathbf{R}$ is said to be* **Lipschitz** *if*

$$\mid f(x) - f(y) \mid \leq d(x, y) \ \forall x, y \in V(G)$$

where d is the graph distance.

2. If f is a random variable, a **Levy mean** *for f (which need not be unique) is a number M_f such that*

$$P\{f \geq M_f\} \geq \frac{1}{2} \text{ and } P\{f \leq M_f\} \geq \frac{1}{2}.$$

Theorem 6.21 *Let G be a graph such that whenever f is a Lipschitz function on G with Levy mean M_f, we have $P\{| f - M_f | > t\} < \alpha$. Then for any subset A of $V(G)$ with $P\{A\} \geq 1/2$, we have, letting $A_{(t)}$ denote the set of points at distance at most t from A in the graph metric that $P\{A_{(t)}\} \geq (1-\alpha)$.*

Proof. [Le], Theorem 6. •

So if we have a worse isoperimetric inequality, the concentration of these functions cannot be as tight as it would have been in the classical model. Another simple reason why various invariants are less tightly concentrated in our models is that usually there will be a slight imbalance between the numbers of reds and of blues, so that the value of the invariant on the reds and on the blues are slightly different (despite the fact that red-red and blue-blue edges arise with the same probability). This ties up with our remarks in Chapter 5 on the failure of the martingale inequalities in our models.

7 Complete graphs and cliques

7.1 Introduction

In this chapter we discuss the closely related subjects of cliques, chromatic numbers and independent sets in our models. Here again there is an extensive theory for the classical model, and many of the results have been sharpened substantially since the appearance of [B]; see [B1] for a summary of some of the progress, much of which is concerned with obtaining tighter concentration results by the use of martingale inequalities. We recall the basic definitions;

Definition 7.1 *A* **complete subgraph** *of order r in a graph G is $C \subset V$, with $\mid C \mid = r$ with all $r(r-1)/2$ edges between these r vertices present in G.*

A **clique** *is a complete subgraph of G which is maximal with respect to the inclusion partial order on subsets of V, so that it is contained in no other complete subgraph.*

The **clique number** $\omega(G)$ *of a graph G is the order of the largest clique in it.*

There is no uniform terminology in the literature; some authors use the term clique for what we have described as a complete graph. Cliques arise in many applications, including design of sequential logic networks in electrical engineering, Bayesian statistics for expert systems in medicine and artificial intelligence, and in taxonomy. There is also a link with ESS theory via the following result;

Theorem 7.1 *Suppose A is an n by n real symmetric matrix with a_{ij}=0 for i=j, otherwise $a_{ij} = \pm 1$. Define a labelled graph $G(A)$ on $\{1, 2, ..n\}$ with an edge between i and j if and only if a_{ij}=1. Then there is a bijection between the set of supports of ESSs of A and the set of cliques of the graph $G(A)$.*

Proof. See [CV]. •

There is thus considerable interest in finding all cliques of a graph. Whilst some effort has been made to get good algorithms for the problem, it is intrinsically difficult; for example, ascertaining whether a graph on n vertices has a clique of order $\geq k$ is an NP-complete decision problem. (A decision problem is a question to which the answer is "yes" or "no"; an instance of it is any object to which the question is addressed; a decision problem is in NP if, given an instance for which the answer is "yes", there is a certificate

verifying the fact which can be checked in polynomial time; in the case of the clique for example, by checking that all edges between the $r \geq k$ vertices are present, and that no other vertex in the graph is adjacent to all r vertices. A subclass of the NP-problems are the problems in P, which is the class of decision problems for which there exists a polynomial time algorithm which solves **every** instance of the problem in polynomial time. It is not known if NP is a larger class than P, but the two classes are widely believed not to coincide. The NP complete problems are a subclass of the NP problems; if any one NP complete problem could be solved in polynomial time, then every problem in NP could be solved in polynomial time. See [ShTa] for more on this; for our purposes, we need only note that it means that the problem is hard). Thus it is of interest to study random behaviour. We summarise the basic facts about the likely numbers and orders of cliques in the classical model;

Theorem 7.2 *In G_α, for α constant and $b = 1/\alpha$, given $0 < \epsilon < 1/2$ a.e. G_α has no clique of order less than $(1-\epsilon)\log_b(n)$ or greater than $(2+\epsilon)\log_b(n)$, but at least one clique of each order r between $(1+\epsilon)\log_b(n)$ and $(2-\epsilon)\log_b(n)$.*

Proof. [B] Theorem XI.4. •

Dual in some sense to the notions of complete graph and clique is the notion of an independent set;

Definition 7.2 *A* **independent set** *of order r in a graph G is a set of r vertices where none of the $r(r-1)/2$ possible edges amongst those r vertices are present. The* **independence number** *$i(G)$ of G is the order of the largest independent set in G.*

Since a complete graph in G is an independent set in its complement (the graph which has the same vertices as G and an edge between two vertices if and only if the corresponding edge is not present in G) we see that the probability that some set of vertices form a complete graph in $\Gamma(n,k,\mathbf{s},P)$ is equal to the probability that the same vertices form an independent set in $\Gamma(n,k,\mathbf{s},J-P)$ where as before J is the k by k matrix of ones; this allows us to translate many statements about the random behaviour of cliques into ones about the behaviour of independent sets, and vice versa.

Definition 7.3 *The* **chromatic number** *$\chi(G)$ is the smallest number of colours with which we can assign a colour to each vertex of G in such a way that no two adjacent vertices have the same colour.*

We re-iterate that these **proper colourings** have nothing to do with the notion of colouring which underlies our models.

Theorem 7.3 *In G_α, with α constant, we have, with $d = 1/(1-\alpha)$,*

$$\chi(G_\alpha) = \left(\frac{1}{2} + o(1)\right) \frac{n}{log_d(n)} \text{ for a.e. } G_\alpha.$$

Proof. [B1] Chapter 4, Theorem 5. •

It is easy to see that $\chi(G) \geq \omega(G)$; the previous result makes it clear that in general it is substantially larger.

7.2 Expected numbers of cliques and complete graphs

We now consider what can be said about expected numbers of cliques in our models; the method will also give the expected number of complete graphs and independent sets. Let the random variables X_r, Y_r and Z_r be the number of cliques of order r, the number of complete graphs of order r and the number of independent sets of order r respectively; if we need to make it clear in which model, we will do so.

Theorem 7.4 *In $\Gamma(n, k, \mathbf{s}, P)$,*

$$\mathbf{E}(X_r) = \binom{n}{r} \sum_{n_1,\ldots n_k} \binom{r}{n_1,\ldots n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{i<j}^{k} p_{ij}^{n_i n_j} \prod_{t=1}^{k} p_{tt}^{\binom{n_t}{2}} \left(1 - \sum_{a=1}^{k} s_a \prod_{b=1}^{k} p_{ab}^{n_b}\right)^{n-r}$$

where the n_i, $1 \leq i \leq k$ satisfy $n_i \geq 0$ and $\sum_{i=1}^{k} n_i = r$, and (with the same notation)

$$\mathbf{E}(Y_r) = \binom{n}{r} \sum_{n_1,\ldots n_k} \binom{r}{n_1,\ldots n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{i<j}^{k} p_{ij}^{n_i n_j} \prod_{t=1}^{k} p_{tt}^{\binom{n_t}{2}}$$

Proof. We give the argument for X_r and then comment briefly on the modification needed for Y_r. First note that by linearity of expectation, considering each r-subset of $V(G)$ separately, the answer is $\binom{n}{r}$ times the probability that r particular vertices form an r-clique. This event in turn consists of these r vertices forming a complete graph and, for each of the other $n-r$ vertices it being untrue that that vertex is joined to all of the r vertices. Observe that whether or not each of these $n-r$ vertices are joined to all the r vertices

happens independently conditional on the colours of the r vertices. Thus, conditioning on the colours of the r vertices, and each of the $n-r$ other vertices, we see that, letting n_i be the number of the r vertices which are of colour i so that $\sum_{i=1}^{k} n_i = r$,

$$\mathbf{E}(X_r) = \binom{n}{r} \sum_{n_1,\ldots n_k} \binom{r}{n_1,\ldots n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{i<j}^{k} p_{ij}^{n_i n_j} \prod_{t=1}^{k} p_{tt}^{\binom{n_t}{2}} \left(1 - \sum_{a=1}^{k} s_a \prod_{b=1}^{k} p_{ab}^{n_b}\right)^{n-r}.$$

For Y_r the only modification required is that we no longer need to worry about whether the other $n-r$ vertices are or are not joined to the complete graph and so the formula becomes

$$\mathbf{E}(Y_r) = \binom{n}{r} \sum_{n_1,\ldots n_k} \binom{r}{n_1,\ldots n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{i<j}^{k} p_{ij}^{n_i n_j} \prod_{t=1}^{k} p_{tt}^{\binom{n_t}{2}}. \bullet$$

Corollary 7.5 *The number of independent sets of order r, Z_r, satisfies*

$$\mathbf{E}(Z_r) = \binom{n}{r} \sum_{n_1,\ldots n_k} \binom{r}{n_1,\ldots n_k} \prod_{l=1}^{k} s_l^{n_l} \prod_{i<j}^{k} (1-p_{ij})^{n_i n_j} \prod_{t=1}^{k} (1-p_{tt})^{\binom{n_t}{2}}. \bullet$$

Corollary 7.6 *In $G_{p,q}$ we have*

$$\mathbf{E}(X_r) = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} p^{r(r-1)/2 - i(r-i)} q^{i(r-i)} \left(1 - \frac{p^i q^{r-i} + q^i p^{r-i}}{2}\right)^{n-r},$$

$$\mathbf{E}(Y_r) = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} p^{r(r-1)/2 - i(r-i)} q^{i(r-i)}$$

and $$\mathbf{E}(Z_r) = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (1-p)^{r(r-1)/2 - i(r-i)} (1-q)^{i(r-i)}.$$

Proof. These formulae are immediate substituting the relevant parameter values into the general formulae in the preceding theorem. •

The above formulae require interpretation if any of the p_{ij} is zero, as then problems about the value to assign to 0^0 arise; we note that the interpretation

$0^0 = 1$ is appropriate here. In particular, in $G_{p,q}$ if $q = 0$ r vertices form a clique if and only if they are all the same colour, all edges amongst themselves exist and for each of the other $n - r$ vertices, not all the edges to the r-set exist, so that

$$\mathbf{E}(X_r) = \binom{n}{r} \frac{p^{r(r-1)/2}}{2^{r-1}} \left(1 - \frac{p^r}{2}\right)^{n-r}$$

and if $p = 0$ so that there are no triangles and so no complete graphs of order ≥ 3, we see that (since any edge is now a 2-clique)

$$\mathbf{E}(X_2) = \frac{qn(n-1)}{4} \text{ and } \mathbf{E}(X_1) = n\left(1 - \frac{q}{2}\right)^{n-1}.$$

We first use this to get an estimate of the actual number of independent sets, using Lemma 5.5.

Theorem 7.7 *In $G_{p,q}$, $\mathbf{E}(Z_r)$ is asymptotically, as r and hence n go to infinity,*

$$\binom{n}{r}(1-p)^{r(r-1)/2} e^{O(r)} \text{ if } p \geq \alpha$$

$$\text{and } \binom{n}{r}\left((1-\alpha)^2 - (p-\alpha)^2\right)^{r(r-1)/4} e^{O(r)} \text{ if } p \leq \alpha.$$

Proof. We have

$$\mathbf{E}(Z_r) = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (1-p)^{r(r-1)/2 - i(r-i)} (1-q)^{i(r-i)}.$$

$$= \binom{n}{r} (1-\alpha)^{r(r-1)/2} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (1+x)^{r(r-1)/2 - i(r-i)} (1-x)^{i(r-i)}.$$

where $x = (\alpha - p)/(1-\alpha)$ has $|\, x \,| < 1$. Consequently we may apply Lemma 5.5 to conclude that

$$\lim_{r\to\infty} \frac{2\log\left(\sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (1+x)^{r(r-1)/2 - i(r-i)} (1-x)^{i(r-i)}\right)}{r(r-1)} = \log(1+x)$$

if $0 < x < 1$, that the limit is zero if $x = 0$ and that it is

$$= \frac{\log(1-x^2)}{2} \text{ if } -1 < x < 0,$$

the error term being $O(1/r)$ in each case. Hence if r goes to infinity with n, we see that

$$\sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (1+x)^{r(r-1)/2-i(r-i)} (1-x)^{i(r-i)} = (1+x)^{r(r-1)/2} e^{O(r)} \text{ if } 0 < x < 1,$$

$$\text{and } \left(1-x^2\right)^{r(r-1)/4} e^{O(r)} \text{ if } -1 < x < 0$$

and is 1 if $x = 0$. Hence the expected number of independent sets of order r in $G_{p,q}$ if r goes to infinity with n is

$$\binom{n}{r} (1-\alpha)^{r(r-1)/2} \left(1 + \frac{\alpha - p}{1-\alpha}\right)^{r(r-1)/2} e^{O(r)} \text{ if } x \in (0,1) \Leftrightarrow p > \alpha,$$

$$\binom{n}{r} (1-\alpha)^{r(r-1)/2} (1 - \left(\frac{\alpha - p}{1-\alpha}\right)^2)^{r(r-1)/4} e^{O(r)} \text{ if } x \in (-1,0) \Leftrightarrow p < \alpha$$

and is of course equal to its classical value otherwise. This gives all the claims on simplifying the formulae. •

Note that, as in Lemma 5.5, the implied constant in the O statements will depend on p as well as α.

Corollary 7.8 *The expected number of complete graphs on r vertices in $G_{p,q}$, as r goes to infinity with n, is*

$$\binom{n}{r} p^{r(r-1)/2} e^{O(r)} \text{ if } p \geq \alpha \text{ and } \binom{n}{r} \left(\alpha^2 - (\alpha - p)^2\right)^{r(r-1)/4} e^{O(r)} \text{ if } p \leq \alpha. \bullet$$

For $p > \alpha$ the error term in Corollary 7.8 always reduces the expectation, as the probability of the complete graph is clearly bounded above by its probability in G_p.

It is worth considering the implications of Corollary 7.8 for the situation where we consider the probability of some sequence of events B_n, where each B_n is that some collection of edges containing a complete graph of order $g(n)$, where $\lim_{n\to\infty} g(n) = \infty$, arise. The corollary says that for large n this probability is close to the probability of B in G_p, in the sense that

$$\lim_{n\to\infty} \left(\frac{P_{p,q}\{B\}}{P_p\{B\}}\right)^{\frac{1}{n^2}} = 1$$

so the bound $P_{p,q}\{B_n\} \leq P_p\{B_n\}$ is less crude than one might imagine.

Another case where we can say something about the probability that some set of n vertices form a complete graph is when both p and q are of order n^{-2}. We shall require again the slight generalisation of the usual product formula for the exponential in Lemma 6.9.

Theorem 7.9 *Suppose $p = c/n^2$, $q = d/n^2$. Then*

$$\lim_{n\to\infty} P\{n \text{ vertices form an independent set}\} = e^{-\frac{c+d}{4}}.$$

In particular the limit depends only on $c+d$ rather than c and d individually.

Proof. The expression to be evaluated is

$$\lim_{n\to\infty} \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} (1-p)^{n(n-1)/2 - i(n-i)} (1-q)^{i(n-i)}$$

$$= \lim_{n\to\infty} (1-p)^{n(n-1)/2} \lim_{n\to\infty} \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} \left(\frac{1-q}{1-p}\right)^{i(n-i)}$$

(since, as we shall see, both limits exist)

$$= \lim_{n\to\infty} (1-p)^{n(n-1)/2} \lim_{n\to\infty} \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} \left(1 + \frac{p-q}{1-p}\right)^{i(n-i)}$$

$$= \lim_{n\to\infty} \left(1 - \frac{c}{n^2}\right)^{n(n-1)/2} \lim_{n\to\infty} \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} \left(1 + \frac{c-d}{n^2(1-\frac{c}{n^2})}\right)^{i(n-i)}.$$

By the lemma, the the first limit is $e^{-\frac{c}{2}}$. As in the proof of 6.9, the second limit is $e^{\frac{c-d}{4}}$, and the result follows multiplying the two limits together. •.

The result for p and q constant above suggests that the number of complete graphs of order r looks, ignoring the small order term, like a monotone increasing function of p (in $G_{p,q}$ subject to fixed α), for r going to infinity with n. It is natural to ask if more generally the number of complete graphs of order r is a non-decreasing function of p. (The case of $r = 2$, when the expected number of complete graphs of order 2 is the expected number of edges, which by linearity of expectation is a constant function of p, makes it clear that it is not always a strictly increasing function). Note that this claim is consistent with the fact that the complete graph is more likely to

arise than classically if $p > q$ (which follows from Theorem 3.1) and is less likely to arise than classically for p in some neighbourhood $(\alpha - \epsilon, \alpha)$ if $p < q$ (by Theorem 3.7, since the newgirth of several edges put together to make a complete graph is of course 3). To investigate the question, we start with the formula in Corollary 7.5 for $\mathbf{E}Y_r$ in $G_{p,q}$. Setting $x = (p - \alpha)/\alpha$, and taking out the term in $\alpha^{n(n-1)/2}$ we see that it is enough to show that the function

$$f_n(x) = \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} (1+x)^{n(n-1)/2 - i(n-i)} (1-x)^{i(n-i)} \quad (*)$$

is an increasing function of x for $x \in [-1, 1]$.

Initial investigations using the computer to expand out the right-hand side of the formula for $f_n(x)$ for several values of n, as a polynomial in x, $\sum_{l=0}^{n(n-1)/2} a_l x^l$ suggests that the coefficients a_l are always non-negative integers; if this is true, it would imply the function is increasing for $x \in [0, 1]$. It also seems that if the large power of $1 + x$ dividing f_n is factored out, the coefficients of the quotient have alternating signs, which would imply that f_n' has no roots in $(-1, 0]$ so has the same sign in the whole interval; since $f_n(-1) = 0$ and $f_n(0) = 1$ that sign must be positive, so f_n is increasing on $[-1, 0]$ also. However proving this seems harder. Note that

$$a_l = \sum_{i=0}^{n} \sum_{j=0}^{l} \frac{\binom{n}{i}}{2^n} \binom{n(n-1)/2 - i(n-i)}{j} (-1)^{l-j} \binom{i(n-i)}{l-j}$$

(where again any $\binom{n}{m}$ where $n < m$ or either n or m is negative, is taken to be zero). It may be worth noting that not all of the summands in the summation are integers. It seems likely that the a_n are in fact counting some quantity, but we do not see how to prove this.

Note that in any interval (a, b) where f_n is convex and $f_n^{(1)}(a) \geq 0$, f is increasing; for convexity implies that $f_n^{(2)}(x) \geq 0 \forall x \in [a, b]$ so $f_n^{(1)}(x)$ is an increasing function; since $f_n^{(1)}(a) \geq 0$, this shows that $f_n^{(1)}(x) \geq 0 \forall x \in [a, b]$ and so f_n is increasing on that interval. However the function is not always convex. Indeed for any n, as $a_2 = 0$ by essentially the same calculation as shows $\mathrm{Var}\mathcal{E}_{p,q} = \mathrm{Var}\mathcal{E}_\alpha$, we have $f_n^{(2)}(x) = n(n-1)(n-2)x + \ldots$ which is < 0 for $x \in (-\epsilon, 0)$ for some $\epsilon > 0$ so f_n is not convex there. In fact our computer calculations suggest that $f_n^{(2)} > 0$ except in some interval $(a_n, 0)$ where $a_n < 0$ and $a_n \to 0$, but again this seems harder to prove.

Note that for n odd the polynomials are symmetric, that is if we write them as $f_n = \sum_{i=0}^{n(n-1)/2} a_i x^i$, $a_{n(n-1)/2-i} = a_i$. For indeed this is equivalent to (as is easily checked) $x^{n(n-1)/2} f(1/x) = f(x)$ for all non-zero x, and it is easy to check that this holds (using the fact that, as n is odd, $i(n-i)$ is even for all values of i, so that $(-1)^{i(n-i)} = 1$ in all cases).

7.3 Clique, chromatic and independence numbers

We now try to obtain at least an estimate of the variance of the number of complete graphs of order r. We are motivated by the notion of using method of moments ([B], page 4) to estimate the probability that there is at least one complete graph of order r; this method is powerful classically, see [B, section XI.1]). We have just seen that

$$\mathbf{E}X_r = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} p^{r(r-1)/2 - i(r-i)} q^{i(r-i)}.$$

Now suppose we have two r-subsets of $V(G)$ which have s vertices in common; note that there are $\binom{n}{r}\binom{r}{s}\binom{n-r}{r-s}$ such pairs of r-subsets. If we condition on there being i reds amongst the s common vertices, j reds in the $r-s$ vertices which are only in one of the two sets and k reds in the $r-s$ vertices which are only in the second set, we see that the number of red-red edges (all of which of course must arise for us to get two complete graphs) is

$$(j(j-1) + i(i-1) + k(k-1))/2 + ij + ik,$$

the number of blue-blue edges is

$$((r-s-j)(r-s-j-1) + (s-i)(s-i-1) + (r-s-k)(r-s-k-1))/2$$
$$+(r-s-j)(s-i) + (r-s-k)(s-i),$$

and the number of same-same edges is thus the sum of the above two numbers

$$f(i,j,k,r,s) = j^2 + k^2 + i^2 + 2ij + 2ik + r^2 - rj - \frac{s^2}{2} - r + \frac{s}{2} + si - rk - 2ir$$

and hence we get that $\mathbf{E}(X_r^2)$ is given by

$$\sum_{s=0}^{r} \sum_{i=0}^{s} \sum_{j=0}^{r-s} \sum_{k=0}^{r-s} q^{n(n-1)/2} \left(\frac{p}{q}\right)^{f(i,j,k,r,s)} \binom{s}{i}\binom{r-s}{j}\binom{r-s}{k}\binom{n}{r}\binom{r}{s}\binom{n-r}{r-s}\frac{1}{2^{2r-s}}.$$

The messy nature of this expression hardly needs emphasising!

Since the numbers of complete graphs of different orders are clearly correlated (for example, the existence of a complete graph on r vertices implies the existence of at least r complete graphs on $r-1$ vertices) it would be harder to get a good grip on the variability of the total number of cliques. However it seems intuitively likely that in the cases where the number of cliques has a multimodal distribution - which, as we shall see shortly, is quite common in our models - the variance of the total number of cliques will be higher than classically.

We close this section with some remarks on the clique number. In G_α $\omega(G)$ is almost determined;

Theorem 7.10 *Given natural numbers $n \geq r$, let r_0 be the positive real number such that*

$$\frac{n^{n+1/2}p^{r_0(r_0-1)/2}}{\sqrt{2\pi}(n-r_0)^{n-r_0+1/2}r_0^{r_0+1/2}} = 1$$

(note the left-hand side is the expression for $\mathbf{E}X_{r_0}$ with the factorial replaced by its Stirling approximation). Then, for a.e. G_α there is a constant $m_0(G)$ such that for $n \geq m_0(G)$,

$$\lfloor r_0(G) - 2\frac{\log_b \log_b(n)}{\log_b(n)} \rfloor \leq \omega(G) \leq \lfloor r_0(G) + 2\frac{\log_b \log_b(n)}{\log_b(n)} \rfloor$$

where $b = 1/\alpha$, and

$$\mid \omega(G) - 2\log_b(n) + 2\log_b \log_b(n) - 2\log_b(\frac{e}{2}) - 1 \mid < \frac{3}{2}.$$

Proof. [B] Corollary XI.2. •

We now consider what happens in our case. The typical variation in the number of reds or blues will be $n/2 \pm K\sqrt{n}$, so that, if q is much smaller than p, so that the largest clique order is likely to be determined by the largest red clique and the largest blue clique, we will have

$$\omega(G) \simeq 2\log_b(n/2 \pm K\sqrt{n}) - 2\log_b \log_b(n/2 \pm K\sqrt{n})$$

$$= 2\log_b(n/2) + 2\log_b(1 \pm \frac{K}{\sqrt{n}}) - 2\log_b(\log_b(n/2) + \log_b(1 \pm \frac{K}{\sqrt{n}})$$

$$\simeq 2\log_b(n/2) - 2\log_b \log_b(n/2)$$

where now $b = 1/p$; and so, in this case at least, we again see that the clique number is close to being determined, though it seems unlikely that it will be as tightly determined as in the second inequality in the previous theorem. Of course this approach gives no insight into the case when $q > p$.

We now make a few remarks about the closely related subject of the independence number. Since the largest clique is, for α fixed, asymptotically of order $2\log(n)/\log(1/\alpha)$ in the classical model, the largest independent set is asymptotically of order $2\log(n)/\log(1/(1-\alpha))$, and if p, q are constant and non-zero, this implies, since the probability that r vertices are an independent set is bounded below by $(1-\max\{p,q\})^{r(r-1)/2}$ and above by $(1-\min\{p,q\})^{r(r-1)/2}$ that the size remains of order of magnitude $\log(n)$. However, if we pass to asking about the order of the largest independent set in a random bipartite graph (this is defined in the same way as before, and is still denoted by $i(G)$), the situation changes dramatically; the largest independent set becomes of order a constant times n. That this is the approximate order is clear, since if $p = 0$ there are no red-red edges, so the about $n/2$ reds are such an independent set. This suggests that our models may be a suitable context in which to understand the transition of the size of the largest independent state from being of order about $c_1 \log(n)$ to order about $c_2 n$. However it is not clear how to proceed in detail with this idea.

We next make some remarks on the chromatic number as p and q vary with their sum fixed. If $q = 0, p = 2\alpha$, and n is large, there will be two components of order about $n/2$, which we can colour separately; thus, by Theorem 7.3

$$\chi(G_{p,q}) \approx (\frac{1}{2} + o(1))\frac{n/2}{\log_{1/1-2\alpha}(n/2)}$$

$$\approx (\frac{1}{2} + o(1))\frac{n}{\log_{1/(1-\alpha)}(n/2)}\frac{\log_{1/(1-\alpha)}(n/2)}{2\log_{1/(1-2\alpha)}(n/2)}.$$

Since the first fraction is approximately the chromatic number of G_α we see that the value here divided by the classical value is approximately

$$\frac{\log_{1/(1-\alpha)}(n/2)}{2\log_{1/(1-2\alpha)}(n/2)}$$

which using the formula $\log_a(b) = \log_c(b)/\log_c(a)$ to put all logarithms into the natural base e, and simplifying (recalling that $\log(1/x) = -\log(x)$) is

$$\frac{\log_e(1-2\alpha)}{2\log_e(1-\alpha)} = 1 + \frac{\alpha}{2} + \frac{\alpha^2}{2} + \frac{3\alpha^2}{4} + \ldots$$

which is close to, and slightly greater than, the classical value. On the other hand, if $q \approx 2\alpha$ and $p \approx 0$, the graph is close to being bipartite and so its chromatic number will be close to 2. It would be interesting to know exactly where the maximum value is.

7.4 Asymptotic theory on expected numbers of cliques

In this section we address questions about asymptotic numbers of cliques and choosing parameter values to maximise them.

The argument for how to maximise the expected number of r-cliques as a function of α in the classical model is easy but I do not know a reference. We differentiate $\mathbf{E}X_r = \binom{n}{r}\alpha^{r(r-1)/2}(1-\alpha^r)^{n-r}$ with respect to α; α is a turning point $\Leftrightarrow \alpha^r = (r-1)/(2n-r-1)$; thus there is only one turning point which is a maximum as $f(\alpha) \geq 0$ on $[0,1]$ and is zero at both ends. At this value of α,

$$\mathbf{E}X_r = \binom{n}{r}\left(\frac{r-1}{2n-r-1}\right)^{(r-1)/2}\left(\frac{2n-2r}{2n-r-1}\right)^{n-r}$$

which by Stirling's formula is asymptotically

$$\left(\frac{r-1}{2n-r-1}\right)^{(r-1)/2}\left(\frac{2n-2r}{2n-r-1}\right)^{n-r}\frac{(n/e)^n\sqrt{2\pi n}}{(r/e)^r\sqrt{2\pi r}((n-r)/e)^{n-r}\sqrt{2\pi(n-r)}}$$

$$= \frac{(r-1)^{(r-1)/2}2^{n-r}n^n}{\sqrt{2\pi r(n-r)/n}r^r(2n-r-1)^{n-(r+1)/2}}.$$

In particular, if $n = ar$ for some constant $a \geq 1$, this is

$$\frac{(\frac{r-1}{r})^{(r-1)/2}(\frac{a^a2^{a-1}}{(2a-1)^{a-1/2}})^r\sqrt{\frac{a(2a-1)}{2\pi(a-1)r}}}{(\frac{(2a-1)^{r-1}}{(2a-1)^r})^{(a-1/2)r-1/2}}$$

which by the product formula for the exponential is asymptotically

$$((\frac{a^a2^{a-1}}{(2a-1)^{a-1/2}})^{1/a})^n\sqrt{\frac{a(2a-1)}{2\pi(a-1)r}}$$

maximised for $((a^a2^{a-1})/((2a-1)^{a-1/2}))^{1/a}$ maximal; its logarithm is $\log(a)+(1-1/a)\log(2)-(1-1/2a)\log(2a-1)$ which has its unique turning point at $a = 2.5$, when the expected number of r-cliques is about $1.6287(1.25)^n/\sqrt{n}$.

A somewhat trickier calculation [CV] shows that the expected total number of cliques $\mathbf{E}(X) = \sum_{r=0}^{n} \mathbf{E}(X_r)$ satisfies

$$\mathbf{E}(X) \sim \sqrt{n} \int_{x=0}^{x=1} \frac{e^{x\gamma/2}}{\sqrt{2\pi x(1-x)}} \left(\frac{e^{-\gamma x^2}(1-e^{-\gamma x})}{x^x(1-x)^{1-x}} \right)^n dx$$

where $\gamma = -n\log(\alpha)$ which is maximised for $\alpha = 2^{-5/n}$, when it is about $1.126(1.25)^n$. For this value of α, the maximal value of $\mathbf{E}(X_r)$ is asymptotically the same as that obtained in the previous paragraph, when we set out only to maximise $\mathbf{E}(X_r)$.

We now consider what can be said in $G_{p,q}$. We again start by considering the case of constant p and q.We have

$$\mathbf{E}(X_r) \text{ in } G_{p,q} = \binom{n}{r} \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} p^{r(r-1)/2-i(r-i)} q^{i(r-i)} (1 - \frac{p^i q^{r-i} + q^i p^{r-i}}{2})^{n-r}.$$

Putting $q = 2a - p$ and expanding about $p = a$ on the computer, we see that the coefficients in $(p-\alpha)$ and $(p-\alpha)^2$ are zero, and that the one in $(p-\alpha)^3$ is

$$\frac{r(r-1)\binom{n}{r}(1-a^r)^{n-r-1}a^{r(r-1)/2}(a^r(2r-3n+2)+r-2)}{6a^3}$$

and consequently

Theorem 7.11 *In $G_{p,q}$ with p and q constant, $p = \alpha$ is a point of increasing or decreasing inflexion according as $a^r(2r-3n+2)+r-2$ is positive or negative.* •

In particular, if $r \leq (3/2 - \epsilon)n$ for some $\epsilon > 0$ when n is sufficiently large, $p = \alpha$ is a point of decreasing inflexion.

Another case in which we can say something is when we put

$$p = \frac{c}{(n-r)^{1/r}} \text{ and } q = \frac{d}{(n-r)^{1/r}}.$$

Then, for fixed r, using the product formula for the exponential for each value of i, we have

$$\lim_{n\to\infty} \frac{(n-r)^{(r-1)/2}\mathbf{E}X_r}{\binom{n}{r}} = \sum_{i=0}^{r} \frac{c^{r(r-1)/2-i(r-i)} d^{i(r-i)} \binom{r}{i} e^{-\frac{c^i d^{r-i} + d^i c^{r-i}}{2}}}{2^r}.$$

Thus we put $d = 2a - c$ and look at the ratio of this to its classical value

$$g_r(c,a) = \lim_{n\to\infty} \frac{\mathbf{E}X_r \text{ in } G_{p,q}}{\mathbf{E}X_r \text{ in } G_a}$$

$$= \sum_{i=0}^{r} \frac{\binom{r}{i}}{2^r} (e^a)^r e^{-(c^i(2a-c)^{r-i}+c^{r-i}(2a-c)^i)/2} (\frac{2a-c}{a})^{i(r-i)} (\frac{c}{a})^{r(r-1)/2-i(r-i)}.$$

For $r = 1$, this expression is equal to 1 clearly, so that the expected number of isolated vertices is (in the limit) independent of c; of course this can be proved, without taking limits, as in Chapter 6. For $r = 2$, we have

$$g_2(c,a) = e^{a^2}(e^{-(c^2+(2a-c)^2)/2}(c/2a) + e^{-c(2a-c)}e^{a^2/2}(2a-c)/2a)/2$$

$$= (e^{-(c-a)^2/2}(c/2a) + e^{(c-a)^2/2}(2a-c)/2a)/2$$

which is $e^{a^2/2}$ if $c = 0$ and $e^{-a^2/2}$ if $c = 2a$. We might well expect that this is a decreasing function of c. To prove this, note that

$$\frac{dg_2}{dc} = \frac{e^{-(c-a)^2}(1 - 2c^2 + 2ac) + e^{(c-a)^2}(6ac - 4a^2 - 2c^2 - 1)}{8a}$$

$$= 0 \Leftrightarrow \frac{2c^2 - 2ac - 1}{6ac - 4a^2 - 2c^2 - 1} = e^{2(a-c)^2} \; (*)$$

which certainly holds for $c = a$; to see that there are no other solutions, note that the numerator on the left-hand side of $(*)$ is less than zero if and only if

$$c \in (\frac{a - \sqrt{a^2+2}}{2}, \frac{a + \sqrt{a^2+2}}{2})$$

but this covers all values of c of interest as $c \in (0, 2a)$; thus, as the right-hand side of $(*)$ is ≥ 1, we thus must have

$$-1 - 2ac + 2c^2 \leq 6ac - 4a^2 - 2c^2 - 1 \Leftrightarrow 4c^2 - 8ac + 4a^2 \leq 0$$

which implies that $c = a$ as before. We next show that $f^{(2)}$ is zero at $c = a$ so this is a point of inflexion; we check (on the computer)

$$\frac{d^2g_2}{dc^2} = \frac{(e^{-(c-a)^2}(4a - 8ac + 4a^2c - 6c + 4c^3)}{8a}$$

$$+ \frac{e^{(c-a)^2}(-4c^3 + 16ac^2 - c(20a^2 + 6) + 8(a^3 + a))}{8a}$$

so if $(*)$ holds we can substitute for $e^{2(c-a)^2}$ and thus

$$\frac{-4e^{-(c-a)^2}(a-c)(8ca^3-20c^2a^2+6a^2+16ac^3-6ac+4c^4+3)}{2a(6ac-4a^2-2c^2-1)}$$

which of course is again 0 at $c=a$. It is also easy to check in the same way that

$$g_2^{(3)}\,|_{c=a}=\frac{-3}{a}<0$$

and so $c=a$ is a point of decreasing inflexion.

The next case is $r=3$, when

$$g_3(c,a)=\frac{c^3e^{-3a(a-c)^2}+3(c^3-4ac^2+4a^2c)e^{a(a-c)^2}}{4a^3}$$

is 0 at $c=0$, 1 at $c=a$ and $2e^{-3a^3}$ if $c=2a$ so has at least one maximum in $(0,2a)$ if $a>(\frac{\log(2)}{3})^{1/3}=0.613623....$ We can compute the first few derivatives at $c=a$ (on the computer) and deduce that

$$g_3(c,a)=1-\frac{(3a^3-1)}{a^3}(c-a)^3+\frac{3(a^3-2)}{2a}(c-a)^4+...$$

so $c=a$ is a point of increasing inflexion if $a<(1/3)^{1/3}$, and of decreasing inflexion if $a>(1/3)^{1/3}$; if $a=(1/3)^{1/3}$ the fourth derivative is negative and so it is a maximum. Since $c=a$ is only a point of inflexion in general, the actual maximum must be elsewhere; use of the implicit plotter in MAPLE suggests, that, for small values of a there may be several values of c which give a turning point, and it is often not too hard to get the computer to give a solution $c\neq a$ for a particular value of a but it seems harder to comment mathematically on these roots.

7.5 Bimodality of the expected number of cliques

We know that, in the classical model, for given α, there is an essentially unique value of r which maximises $\mathbf{E}(X_r)$. Indeed

$$\frac{\mathbf{E}(X_{r+1})}{\mathbf{E}(X_r)}=\frac{\binom{n}{r+1}\alpha^{(r+1)r/2}(1-\alpha^{r+1})^{n-r-1}}{\binom{n}{r}\alpha^{(r-1)r/2}(1-\alpha^r)^{n-r}}=\frac{(n-r)\alpha^r(1-\alpha^{r+1})^{n-r-1}}{(r+1)(1-\alpha^r)^{n-r}}$$

which is a strictly decreasing function of r, and so either there is some value of r for which $\mathbf{E}(X_r) = \mathbf{E}(X_{r+1})$ are the two maximum values of the expectation (for example, if $n = 3$, $\alpha = 1/\sqrt{2}$, $\mathbf{E}(X_1) = \mathbf{E}(X_2)$), or the maximum is unique. [CV] shows that if α is constant, this modal clique order is asymptotically $\log_b(n) - \log_b \log_b \log_b(n)$ where $b = 1/\alpha$. In fact of course, in the classical model, we also know that the distribution of clique orders is tightly concentrated around the mean.

In our models, however, numerical investigations rapidly make it clear that $\mathbf{E}(X_r)$ often has several maxima. The idea is clear; in G_α a.e. graph has complete graphs of all orders less than $\omega(G) \simeq 2\log_b(n)$. Now with probability tending to 1 there are $n(1 + o(n^\kappa))/2$ reds and blues, for any $\kappa \in (-1/2, 0]$, and so the largest complete graphs in the reds or the blues are of order about $2\log_{1/p}(n)$; the argument is now that if q is sufficiently small, the prospects of a multicoloured clique with large numbers of vertices of both colours, of order near $2\log_{1/p} n$ are very small. However a precise result on just how unlikely this is seems rather harder to come by.

Note also that our computations suggest that with several colours it is possible to have several modes.

7.6 The evolving clique in $G_{p,q}$.

Another approach considers a given clique evolving in time. (For another kind of evolution, see [W]). In the classical form of the evolution we consider, we start at time 1 with one vertex, and expand to a clique of order 2 at time 2 with probability α, else staying at 1; if we eventually get to a clique of order 2, we then might get to a clique of order 3 with probability α^2 at each trial, etc. Thus the evolution of X_n, the order of the clique of the **initial** vertex (we are not interested in any other cliques) is a discrete-time pure birth process with $P\{X_{n+1} = r + 1 \mid X_n = r\} = \alpha^r$ and $P\{X_{n+1} = r \mid X_n = r\} = 1 - \alpha^r$. This is in many ways a more realistic model for the applications to ESSs, since it better reflects the way that, in reality, strategies are added to those already known. This process was studied by Cannings and Vickers [CV]. The main problem, of understanding the distribution of X_n, is rather intractable; indeed, if N_j is the time spent in state j, clearly $P\{N_j = i\} = (1-\alpha^j)^{i-1}\alpha^j$ for $i \geq 1$ and the N_j are independent so $P\{X_n = r\}$ is the sum over all partitions of n into r strictly positive integers of the $P\{N_1 = i_1, \ldots N_{r-1} = i_{r-1}, N_r \geq i_r\}$.

This probability is

$$\prod_{j=1}^{r-1}((1-\alpha^j)^{i_j-1}\alpha^j)(1-\alpha^r)^{i_r-1} = \alpha^{r(r-1)/2}\prod_{j=1}^{r}(1-\alpha^j)^{i_j-1}$$

$$\Rightarrow P\{X_n = r\} = \sum \alpha^{r(r-1)/2}\prod_{j=1}^{r}(1-\alpha^j)^{i_j-1}$$

the summation being over

$$\{i_1, ...i_r \text{ such that } i_j \geq 1 \forall 1 \leq j \leq r \text{ and } i_1 + ... + i_r = n.\}$$

and this expression is is not too easy to work with. Thus attention switches to asymptotic results on $\mathbf{E}(X_n)$; these are also discussed in [CV] where the fact that, for large n, $\mathbf{E}(X_n) \sim \log_b n$ where $b = \alpha^{-1}$ is stated (without proof) to follow from results of Grimmett and McDiarmid [GM] on the clique number of random graphs in G_α. In fact the result can be obtained more easily, with some information about the error; we sketch the argument.

Lemma 7.12 *For any $\epsilon > 0$, $P\{X_n \geq n^{1/2+\epsilon}\}$ is exponentially small in some positive power of n for large enough n.*

Proof. The probability is clearly (writing $m = n^{1/2+\epsilon}$) $\leq \binom{n}{m}2^{-m(m-1)/2}$ since $m(m-1)/2$ edges must have formed. Now as $n - m(m-1)/2$ is about $-n^\delta$ for some $\delta > 0$ for n large, and the binomial coefficient is less than 2^n, the result follows. •

The estimate is very crude and could probably be improved.

Theorem 7.13 $\mathbf{E}(X_n) \sim \log_b(n)$.

Proof. Let Y_r be the time at which we enter state $r+1$. Thus $Y_r = \sum_{i=1}^{r} N_i$, where N_i is (as before) the time spent in state i; the N_i are geometric with parameter α^i and so mean $\alpha^{-i} - 1$. Thus

$$\mathbf{E}(Y_r) = \sum_{i=1}^{r}(\alpha^{-i} - 1) = \frac{1-\alpha^{r+1}}{\alpha^r - \alpha^{r+1}} - r.$$

Now $\mathbf{E}(X_n)$ is that r such that

$$\mathbf{E}(Y_r) \leq n < \mathbf{E}(Y_{r+1})\,(*).$$

The inequality on the right thus gives that $(n+r)(\alpha^r - \alpha^{r+1}) \leq 1 - \alpha^{r+1}$; this implies that $b^r \geq (n+r)(1-\alpha)+\alpha$ and so that $r \geq log_b(n+r)+\log_b(1-\alpha+\alpha/(n+r))$. Since $r < n^{1/2+\epsilon}$ with probability tending to 1 by the lemma, $\log_b(n+r) = \log_b(n)$ plus a smaller order error term. Similarily, the lower bound in $(*)$ shows that $r - 1 \leq \log_b(n+r+1)$, and again by the lemma $\log_b(n)$ is close to r. •

The argument also suggests that the variability of X_n for large n should be small; this is supported by simulations.

We next consider the version of this process analogous to $G_{p,q}$ where each vertex, having been thrown down, is randomly coloured red or blue, and then same-same edges arise with probability p and red-blue edges with probability q, so that states of the system are pairs (i,j) with i is the number of reds and j the number of blues, though we shall sometimes still use the notation $P\{X_n = r\}$ for the probability that the total number of vertices is r if this is not confusing. Then, in ${}_sG_{p,q,r}$,

$$P\{X_n = (i,j) \mid X_{n-1} = (i,j)\} = (1-p^iq^j)s + (1-q^ir^j)(1-s)$$

$$P\{X_n = (i,j+1) \mid X_{n-1} = (i,j)\} = (1-s)q^ir^j$$

$$P\{X_n = (i+1,j) \mid X_{n-1} = (i,j)\} = sp^iq^j$$

Setting $F_n(x,y) = \sum_{1\leq i,j\leq n} P\{X_n = (i,j)\}x^iy^j$, so that

$$P\{X_n = r\} = \sum_{i=0}^{r} \frac{1}{i!(r-i)!} \frac{\delta^r F_n}{\delta x^i \delta y^{r-i}} \mid_{x=y=0},$$

we see that the polynomials $F_n(x,y)$ satisfy the recurrence

$$F_n(x,y) = F_{n-1}(x,y) + s(x-1)F_{n-1}(px,qy) + (1-s)(y-1)F_{n-1}(qx,ry)$$

with $F_1(x,y) = sx + (1-s)y$; however, it does not seem obvious how to get even good estimates of the solutions of these equations.

Whilst it is obvious that $\log_c(n) \leq \mathbf{E}(X_n) \leq \log_d(n)$ where $c = 1/\min\{p,q\}$ and $d = 1/\max\{p,q\}$, it is not clear where in this range it will be in general. However if $q = 0$ and $p = 2\alpha$ (so $\alpha \leq 1/2$) the evolving clique will be monochrome so we would expect X_n to be roughly the same as in the classical evolving clique with $n/2$ vertices and probability 2α, whence

$$\mathbf{E}X_n \simeq \log_{1/2\alpha}(n/2) = \frac{\log_{1/\alpha}(n) - \log_{1/\alpha}(2)}{1 + \log_{1/\alpha}(1/2)};$$

since $\alpha \leq 1/2$, $1 + \log_{1\alpha}(1/2) \in (0,1)$ so $\mathbf{E}X_n$ is larger than classically. If $p = 0$ the largest clique has order 2. This suggests that the order is an increasing function of p for fixed α.

One feature that emerged from simple simulation experiments is that there is a tendency, when $p > q$, for only a few vertices of the opposite colour to the initial vertex to be in the clique; we will call such a vertex entering the clique an **infiltration**. On the other hand, if there is infiltration, the growth rate of the clique is reduced substantially; this suggests that the variability will be much larger in our models. We are thus lead to study the number of infiltrations. We start by considering the event A_n, that no vertex of the opposite colour to the initial vertex is absorbed into the clique whilst it has n or fewer vertices, and their intersection A.

Theorem 7.14 *In $G_{p,q}$, if $p > q$ $P\{A\} > 0$, but if $p \leq q$ $P\{A\} = 0$.*

Proof. Let $B_{i,j}$ be the event that no infiltration occurs when the clique has i vertices of the initial colour, and j of the opposite colour. Then

$$P\{B_{i,j}\} = \sum_{r=1}^{\infty} P\{B_{i,j} \cap \text{we leave state } (i,j) \text{ after } r \text{ trials}\}.$$

Of course $B_{i,j}$ and a vertex being absorbed on the rth trial during that state is the event that a vertex of the initial colour is absorbed on the rth trial. In each such trial, independently, the probability that the vertex is the initial colour is $1/2$ and the probability that such a vertex is absorbed is $p^i q^j$; thus

$$P\{B_{i,j}\} = \sum_{r=1}^{\infty} \frac{p^i q^j}{2}(1 - \frac{p^i q^j + q^i p^j}{2})^{r-1} = \frac{p^i q^j}{p^i q^j + q^i p^j} = \frac{1}{1 + (p/q)^{j-i}}.$$

$$\Rightarrow P\{A_i\} = \prod_{j=1}^{i} P\{B_{j,0}\} = \prod_{k=1}^{i} \frac{p^k}{p^k + q^k} = \prod_{k=1}^{i}(1 - \frac{(q/p)^k}{1 + (q/p)^k})$$

$$\Rightarrow P\{A\} = \prod_{k=1}^{\infty}(1 - \frac{(q/p)^k}{1 + (q/p)^k}) \geq \prod_{k=1}^{i}(1 - (\frac{q}{p})^k) \text{ as } p > q.$$

However it is well-known that for $0 \leq z < 1$, the function $\prod_{j=1}^{\infty}(1 - z^j)$ is convergent to a non-zero limit (recall the product of the $(1-a_n)$ for $a_n \in [0,1]$ is non-zero if and only if $\sum_{i=1}^{\infty} a_i$ is finite).

On the other hand, if $p \leq q$, then, setting aside the trivial case $p = q = 0$ we have jth factor in the product for $P\{A_i\}$ is

$$\frac{1}{1 + (\frac{q}{p})^j} \leq 1/2 \Rightarrow P\{A\} = \lim_{i \to \infty} P\{A_i\} = 0 \text{ if } p \leq q. \bullet$$

In fact, if $p = q$ by symmetry the number of vertices of the initial colour is $\text{Bin}(n-1, 1/2)$ distribution when there are n vertices in total; and if $q > p$, let S_n be the number, from the n vertices in total, of the initial colour minus the number of the opposite colour; thus S_n is a Markov chain. If $S_n = r$ there are $(n+r)/2$ vertices of the initial colour and $(n-r)/2$ of the opposite colour, and so, using the formulae for $P\{B_{i,j}\}$ and simplifying

$$P\{S_{n+1} = r + 1 \mid S_n = r\} = \frac{(p/q)^{r/2}}{(p/q)^{r/2} + (q/p)^{r/2}}$$

$$P\{S_{n+1} = r - 1 \mid S_n = r\} = \frac{(q/p)^{r/2}}{(p/q)^{r/2} + (q/p)^{r/2}}$$

so (as $q/p > 1$) the chain tend to drift back towards zero when away from it, and in state zero, it is equally likely to go in either direction. This suggests that the chain should settle down, spending most of its time close to state zero; to formalise this, we show it has a stationary distribution whose terms are exponentially small in n for large n. Indeed the process is a mixture of two birth and death processes, one of them being what happens for $r > 0$, the other for $r < 0$. The standard analysis ([D, p301]) shows that each of these has a stationary distribution, which for $r > 0$ is given by

$$\pi_i = \prod_{k=1}^{i} \frac{P\{S_{n+1} = k \mid S_n = k-1\}}{P\{S_{n+1} = k-1 \mid S_n = k\}}$$

$$= \prod_{k=1}^{i} \left(\frac{(p/q)^{(k-1)/2}}{(p/q)^{(k-1)/2} + (q/p)^{(k-1)/2}} \frac{(q/p)^{k/2} + (p/q)^{k/2}}{(q/p)^{k/2}} \right.$$

$$= \prod_{k=1}^{i} \frac{(p/q)^{(2k-1)/2}((q/p)^{k/2} + (p/q)^{k/2})}{(p/q)^{(k-1)/2} + (q/p)^{(k-1)/2}}$$

Since $q > p$, these terms are indeed small for large values of i, since then the fraction in the product is approximately $(p/q)^{2k-2}$ which is small for

large values of k. The stationary distribution for $r < 0$ is very similar, and the stationary distribution for the whole process is the average of the two stationary distributions.

Thus we now concentrate on the case $p > q$, and let $t = q/p \in [0,1)$. Whilst we have shown $P\{A\} > 0$, our estimate of it is rather crude. This suggests that we might get some understanding by investigating the probabilities of other small numbers of infiltrations, to see if a pattern emerges. We start with the probability of exactly one infiltration.We have, using the above formula for $P\{B_{i,j}\}$, that

$$P\{\text{one infiltration up to state } n\} = \sum_{i=1}^{n} P\{\text{infiltration in state } i \text{ only}\}$$

$$= \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \frac{p^j}{p^j + q^j}\right) \frac{q^i}{p^i + q^i} \prod_{i+1}^{n} \frac{p^j q}{p^j q + q^j p}$$

$$= \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \left(1 - \frac{q^j}{p^j + q^j}\right)\right)\left(1 - \frac{p^i}{p^i + q^i}\right) \prod_{j=i+1}^{n} \left(1 - \frac{q^j p}{p^j q + q^j p}\right) (*).$$

Since this expression is

$$\geq \sum_{i=1}^{n} \left(\prod_{j=1}^{n} \left(1 - \left(\frac{q}{p}\right)^j\right)\right) \frac{1}{(1-(q/p)^i)} \frac{q^i}{p^i + q^i} = \sum_{i=1}^{n} \prod_{j=1}^{n} \left(1 - \left(\frac{q}{p}\right)^j\right) \frac{p^i q^i}{p^{2i} - q^{2i}}$$

which for $p > q$ is

$$\geq \sum_{i=1}^{n} \prod_{j=1}^{n} \left(1 - \left(\frac{q}{p}\right)^j\right) \frac{q^i}{p^i} = \prod_{j=1}^{n} \left(1 - \left(\frac{q}{p}\right)^j\right) \frac{1 - (q/p)^{n+1}}{1 - q/p} = \prod_{j=2}^{n+1} \left(1 - \left(\frac{q}{p}\right)^j\right)$$

it converges as before to a non-zero limit. In fact similar arguments will show that, for any fixed r, the probability that at most r infiltrations occur will, when $q > p$, tend to a non-zero limit.

To evaluate the limit, it seems to be easier to consider $q_i(n)$, the ratio of the probability of exactly i infiltrations when the clique has exactly n vertices to the probability of none, with $q_i = \lim_{n\to\infty} q_i(n)$. Then, setting $t = q/p$, we have

Theorem 7.15

$$q_1(n) = \frac{t(1 + t^{n-1})(1 - t^n)}{2(1 - t)} \quad \textit{and}$$

$$q_2(n) = \frac{t^2(1-t^n)(1-t^{n-1}))(1+t^n)(1+t^{n-3})}{2(1-t^2)^2}.$$

Proof. We have

$$\sum_{i=1}^{n} \frac{q^i}{p^i} \prod_{j=i}^{n-1} \frac{p^j q/(p^j q + q^j p))}{p^{j+1}/(p^{j+1} + q^{j+1})}$$

$$= \sum_{i=1}^{n} \frac{q^i}{p^i} \prod_{j=i}^{n} \frac{q(q^{j+1} + p^{j+1})}{p(pq^j + p^j q)} = \sum_{i=1}^{n} t^i \prod_{j=i}^{n-1} \frac{t(1+t^{j+1})}{t^j + t}$$

$$= \sum_{i=1}^{n} t^i \prod_{j=i}^{n-1} \frac{1+t^{j+1}}{1+t^{j-1}} = (1+t^n)(1+t^{n-1}) \sum_{i=1}^{n} \frac{t^i}{(1+t^{i-1})(1+t^i)}$$

so it suffices to prove by induction that $\forall n \geq 2$

$$\sum_{i=1}^{n} \frac{t^i}{(1+t^{i-1})(1+t^i)} = \frac{t(1-t^n)}{2(1-t)(1+t^n)} \; (*).$$

The case $n = 2$ is a short calculation and if the claim holds for n, then

$$\sum_{i=1}^{n+1} \frac{t^i}{(1+t^{i-1})(1+t^i)} = \frac{t(1-t^n)}{2(1-t)(1+t^n)} + \frac{t^{n+1}}{(1+t^n)(1+t^{n+1})}$$

$$= \frac{t(1-t^n)(1+t^{n+1}) + 2t^{n+1}(1-t)}{2(1-t)(1+t^n)(1+t^{n+1})}$$

and a short calculation shows that the numerator of this last fraction is equal to $t(1+t^n)(1-t^{n+1})$, giving the required result. (In particular, letting $n \to \infty$, we see that $p_1 = t/(2(1-t))$, which is a monotone increasing function of t and is less than 1 if and only if $t < 2/3$). For the second claim, as the probability of two infiltrations up to time n is (writing the expression in terms of t)

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\prod_{k=1}^{i-1} \frac{1}{1+t^k}\right) \frac{t^i}{1+t^i} \left(\prod_{l=i+1}^{j-1} \frac{t}{t+t^{l-1}}\right) \frac{t^{j-1}}{t^{j-1}+t} \prod_{m=j+1}^{n} \frac{t^2}{t^2+t^{m-2}}.$$

$$\Rightarrow p_2(n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} t^i \left(\prod_{l=i+1}^{j-1} \frac{t(1+t^l)}{(t+t^{l-1})}\right) \frac{t^{j-1}(1+t^j)}{t^{j-1}+t} \prod_{m=j+1}^{n} \frac{t^2(1+t^m)}{t^2+t^{m-2}}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} t^i \left(\prod_{l=i+1}^{j-1} \frac{t(1+t^l)}{t+t^{l-1}}\right) \frac{t^{j-1}(1+t^j)}{t^{j-1}+t} \prod_{m=j+1}^{n} \frac{t^2(1+t^m)}{(t^{m-2}+t^2)}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} t^i \Big(\prod_{l=i+1}^{j-1} \frac{1+t^l}{1+t^{l-2}} \Big) \frac{t^{j-2}(1+t^j)}{t^{j-2}+1} \prod_{m=j+1}^{n} \frac{1+t^m}{t^{m-4}+1}$$

$$= \prod_{k=n-3}^{n} (1+t^k) \sum_{i=1}^{n-1} \frac{t^i}{(1+t^{i-1})(1+t^i)} \sum_{j=i+1}^{n} \frac{t^{j-2}}{(1+t^{j-3})(1+t^{j-2})}$$

$$= \prod_{k=n-3}^{n} (1+t^k) \sum_{i=1}^{n-1} \frac{t^i}{(1+t^{i-1})(1+t^i)} \sum_{j=i-1}^{n-2} \frac{t^j}{(1+t^{j-1})(1+t^j)}$$

But we can replace the double sum by (using $(*)$ for $n \geq 2$)

$$\sum_{i=1}^{n-1} \frac{t^i}{(1+t^i)(1+t^{i-1})} \Big(\frac{t(1-t^{n-2})}{2(1+t^{n-2})(1-t)} - \frac{t(1-t^{i-2})}{2(1+t^{i-2})(1-t)} \Big)$$

$$= \frac{t^2(1-t^{n-1})(1-t^{n-2})}{4(1+t^{n-2})(1+t^{n-1})(1-t)^2} - \sum_{i=1}^{n-1} \frac{t^{i+1}(1-t^{i-2})}{2(1+t^i)(1+t^{i-1})(1+t^{i-2})(1-t)}.$$

We prove by induction that, for $n \geq 6$, we have the identity

$$\sum_{i=2}^{n} \frac{t^{i+1}(1-t^{i-2})}{2(1+t^i)(1+t^{i-1})(1+t^{i-2})(1-t)} = \frac{t^4(1-t^{n-1})(1-t^{n-2})}{2(1+t^n)(1+t^{n-1})(1-t^2)^2}$$

The case $n = 6$ is a tedious calculation, and if the relation holds for n,

$$\sum_{i=3}^{n+1} \frac{t^{i+1}(1-t^{i-2})}{2(1+t^i)(1+t^{i-1})(1+t^{i-2})(1-t)}$$

$$= \frac{t^4(1-t^{n-1})(1-t^{n-2})}{2(1+t^n)(1+t^{n-1})(1-t^2)^2} + \frac{t^{n+2}(1-t^{n-1})}{2(1+t^{n+1})(1+t^n)(1+t^{n-1})(1-t)}$$

$$= \frac{t^4(1-t^{n-1})(1-t^{n-2})(1+t^{n+1}) + t^{n+2}(1-t^{n-1})(1-t)(1+t)^2}{2(1+t^n)(1+t^{n-1})(1+t^{n+1})(1-t^2)^2}$$

so the result follows from a short calculation to check the identity

$$t^4(1-t^{n-1})(1-t^{n-2})(1+t^{n+1}) + t^{n+2}(1-t^{n-1})(1-t)(1+t)^2$$

$$= t^4(1-t^n)(1-t^{n-1})(1+t^{n-1}).$$

Consequently we have that the double sum under consideration is

$$\frac{t^2(1-t^{n-2})(1-t^{n-1})}{4(1+t^{n-2})(1-t^{n-1})(1-t)^2} - \frac{t^4(1-t^{n-2})(1-t^{n-3})}{2(1+t^{n-1})(1+t^{n-2})(1-t^2)^2} + \frac{t^2}{4(1+t)^2}$$

the last term being to allow for the $i = 1$ term of the summation; the $i = 2$ term is of course zero. Thus the double sum simplifies further to

$$= \frac{t^2(1-t^n)(1-t^{n-1})}{2(1-t^2)^2(1+t^{n-2})(1+t^{n-1})}.$$

Thus

$$q_2(n) = \frac{t^2(1-t^n)(1-t^{n-1}))(1+t^n)(1+t^{n-3})}{2(1-t^2)^2.}$$

In particular, letting $n \to \infty$

$$q_2 = \frac{t^2}{2(1-t^2)^2}$$

which is again a monotone increasing function of $t \in [0,1]$; $q_2 < 1$ if and only if $2t^4 - 5t^2 + 2 > 0$, which for $t \in [0,1]$ is true when $t < 1/\sqrt{2}$. •

By similar arguments, it is easy to show that $q_3(n)$ is given by

$$\sum t^i \Big(\prod_{l=i+1}^{j-1} \frac{1+t^l}{1+t^{l-2}}\Big)\frac{t^{j-2}(1+t^j)}{t^{j-2}+1}\Big(\prod_{m=j+1}^{k-1} \frac{1+t^m}{t^{m-4}+1}\Big)\frac{t^{k-4}(1+t^k)}{1+t^{k-4}} \prod_{r=k+1}^{n} \frac{1+t^r}{t^{r-6}+1}$$

the sum being a triple sum over the range $1 \le i \le n-2$, $i+1 \le j \le n-1$ and $j+1 \le k \le n$, and it is now obvious how to write down the formula for $q_r(n)$ in principle. However it is not possible to simplify this expression, even for $q_3(n)$, since a certain associated triple sum lacks any obvious simplification. However we have not shown that a simple closed formula does not exist, and there may well be one. Note that one obvious guess, that $q_3(n) = t^3/(2(1-t^3)^3)$, can be shown to be untrue.

One might be tempted to speculate that the probability of only finitely many infiltrations is (for $p > q$) one in the limit. However this will not be true. Indeed if the second vertex is an infiltration we will have one red and one blue; thus if the claim were true, this would with probability one end up with only finitely many reds, and also with only finitely many blues, so would be finite, which is nonsense.

One might also be tempted to try to study $p_i(n)$, the probability that when we have n vertices in total there have been i infiltrations via the obvious relation

$$p_i(n) = \frac{p_{i-1}(n-1)p^{i-1}q^{n-i}}{p^{i-1}q^{n-i}+q^{i-1}p^{n-i}} + \frac{p_i(n-1)p^{n-1-i}q^i}{p^{n-1-i}q^i+q^{n-1-i}p^i}.$$

If we set $F(u,v) = \sum_{i,n} p_i(n) u^i v^n$, and $t = q/p$ as before, then, purely formally, this implies

$$F(u,v) = uv \sum_{j=0}^{\infty} (-1)^j F(ut^{2j}, vt^{-j}) + v \sum_{j=0}^{\infty} (-1)^j F(ut^{-2j}, vt^j)$$

but it is unclear how to solve this functional equation, and there would be serious issues of convergence to deal with.

8 Tournaments with correlation structure

8.1 Introduction.

By a **tournament** T we mean a complete graph where the edge between vertices a and b is oriented $a \to b$ or $b \to a$ (not both). As the name suggests, the simplest way to think about them is in terms of the results of a tournament where each player plays each other with no draws, a win for i against j corresponding to an edge $i \to j$. [M] is a basic reference for material on tournaments. They arise in practice in experimental design where paired comparisons are carried out, with an arrow $i \to j$ if treatment i is preferable to treatment j, and in the study of dominance relations in biology, with an arrow going $i \to j$ if the individual i dominates (in whatever sense) individual j.

As the above two examples suggest, often an interesting problem is trying to determine the real underlying strengths of the players. Since a **cycle**, i.e a sequence of distinct edges $a \to b \to c...z \to a$ shows an apparent inconsistency in the ordering, numbers of cycles are closely tied up with this question (though there are also other arguably more sophisticated measures available, such as the top eigenvector of a suitable adjacency matrix associated with T; see [M, Chapter 18]). (Some authors call what we have called a cycle a directed cycle, using the term cycle for any set of edges which are a cycle in the undirected graph but may not be in the directed one).

The usual notion of random tournament arises when each edge goes $a \to b$ or $b \to a$ equiprobably and independently. Here some reasonable theory has been developed on the distribution of numbers of cycles and the probability that the tournament is **irreducible**, i.e that there is no partition of the vertex set into two non-empty sets A and B with all the edges between A and B going in the same direction, so that a reducible tournament is one with two groups of players, with each member of one group stronger than all members of the other. In fact ([M], page 13), a tournament is irreducible if and only if it has at least one cycle of length n if and only if it has at least one cycle of each length r such that $3 \leq r \leq n$, so there is a close link between reducibility properties and cycles.

In this chapter, we will seek to develop a theory of tournaments where the orientation of the edges depends on the random types of the distinct vertices, and discuss how techniques from previous chapters can be modified to give some insight into the behaviour of the model. The form of inhomogeneity

suggested by the above examples is one where there are subgroups of players of about equal standard who are as likely to win as to lose in matches amongst themselves, but tend to lose or win against other groups. So we have;

Definition 8.1 *An* **RRC tournament model** $\tau(n, k, \mathbf{s}, P)$ *is one where each of n vertices is independently assigned one of k colours, receiving the j-th with probability* s_j *(so* $\mathbf{s} = (s_1, ...s_k) \in \Delta_k$*) and then an edge between a vertex of colour i and colour j goes* $i \to j$ *with probability* p_{ij} *and* $j \to i$ *with probability* p_{ji} *(so that P is a k by k real matrix).*

Two consistency conditions forced on us by the definition are that

$$p_{ii} = \frac{1}{2} \forall i \text{ and } p_{ij} = 1 - p_{ji} \forall i, j.$$

Note that in any $\tau(n, k, \mathbf{s}, P)$ model, the overall probability that an edge goes $i \to j$ and the probability that it goes $j \to i$ are both still equal to $1/2$; the differences from the classical model are due to correlation structure.

In fact, we shall also sometimes consider **partial tournaments**, where some of the edges may not exist. For convenience, we restrict to the case where we continue to insist that all edges between vertices of the same colour arise. Then we have

$$P\{\text{colour } i \to \text{colour } j\} = p_{ij} \text{ and } P\{\text{colour } j \to \text{colour } i\} = p_{ji}$$

where (since we continue to insist that there is at most one edge between any two vertices) we must still have $p_{ij} + p_{ji} \leq 1$, but we do not now insist that $p_{ij} = 1 - p_{ji}$.

Again we shall often be concerned with the case of two colours red and blue, with edges between a red vertex and a blue one going from red to blue with probability p; if vertices are red with probability s we shall denote the model ${}_sT_p$ and if $s = 1/2$ we shall shorten this to T_p. We shall also use the partial tournament model $T_{p,q}$ where $q \leq 1 - p$ is the probability that the edge goes from blue to red.

It is clear that the distributions of most random variables of interest are invariant under the transformation

$$(p, s) \to (1 - p, 1 - s) \text{ in } {}_sT_p.$$

8.2 Probabilities of paths and cycles.

We first obtain the probability of an r-cycle. The arguments are similar to Theorems 2.6 and 2.10; again there is a result for any $\tau(n, k, \mathbf{s}, P)$ which unfortunately is not too explicit, and a more direct argument for the special case of T_p which gives more information. This time we deal with the former case first.

Theorem 8.1 *Define Q by $q_{ij} = \sqrt{s_i} p_{ij} \sqrt{s_j}$ and $\mathbf{v}$ by $v_i = \sqrt{s_i}$ as before. Then the probability of an r-cycle in any model of random partial tournaments is equal to (letting λ_i, $1 \leq i \leq k$ be the eigenvalues of Q) $\sum_{i=1}^{k} \lambda_i^r$.*

If $\mathbf{v}$ is not an eigenvector of Q, the probability of any cycle of sufficiently large length is always less than classically.

In a model of tournaments, if $\mathbf{v}$ is an eigenvector, cycles of length congruent to ± 1 (mod 4) have the same probability as classically, those of length congruent to 2 (mod 4) are less probable, and those of length congruent to 0 (mod 4) are more probable.

Proof. Conditioning on the colours of the r vertices, we have that

$$P\{1 \to 2 \ldots r \to 1\} = \sum_{i_1, \ldots i_r = 1}^{k} s_{i_1} s_{i_2} \ldots s_{i_r} p_{i_1 i_2} p_{i_2 i_3} \ldots p_{i_r i_1}$$

By the definition of Q, this implies

$$P\{1 \to 2 \to \ldots \to r \to 1\} = \sum_{i_1, \ldots i_r = 1}^{k} q_{i_1 i_2} q_{i_2 i_3} \ldots q_{i_r i_1}.$$

and we can now sum out the variables one by one, each such summation corresponding to a matrix multiplication. Doing this, we get

$$\sum_{i}^{k} (Q^r)_{ii} = tr\,(Q^r) = \sum_{i=1}^{k} \lambda_i^r$$

by general matrix theory, as required.

To get the remaining assertion, we deal only with the case when the model is of tournaments rather than partial tournaments, as the probability of a cycle in the partial tournament is always less than or equal to the probability of it in a model of tournaments obtained by adding a little on to some entries so as to have $p_{ij} + p_{ji} = 1$. Then, because we have a tournament, there is a

skew symmetric matrix T such that, letting J be the k by k matrix of ones as before,

$$P = \frac{1}{2}J + T \Rightarrow Q = A + S \text{ where } a_{ij} = \frac{1}{2}\sqrt{s_i s_j} \text{ and } s_{ij} = \sqrt{s_i s_j}t_{ij}$$

so S is still skew symmetric.

We first show that the top eigenvalue of Q is at most $1/2$. We note that A is symmetric, so it has an orthonormal basis of eigenvectors $\mathbf{e}_i$, for $1 \leq i \leq k$. Also A has rank 1. In addition we have that

$$(A\mathbf{v})_i = \sum_{j=1}^{k} \frac{1}{2}\sqrt{s_i s_j}\sqrt{s_j} = \frac{1}{2}\sqrt{s_i} \Rightarrow A\mathbf{v} = \frac{1}{2}\mathbf{v}$$

so $\mathbf{v} = \mathbf{e}_1$ without loss of generality and thus the other $\mathbf{e}_i$, $2 \leq i \leq k$ all have eigenvalue 0.

Now Q is a non-negative matrix so has a non-negative eigenvalue, λ say of maximum modulus; if $\mathbf{w}$ is a corresponding eigenvector, normalised to have modulus 1, we have

$$\lambda = \mathbf{w}^T Q\mathbf{w} = \mathbf{w}^T A\mathbf{w}$$

since as S is skew symmetric $\mathbf{x}^T S\mathbf{x} = 0$ for all vectors $\mathbf{x}$. Now writing $\mathbf{w} = \sum_{j=1}^{k} \kappa_j \mathbf{e}_j$, where the κ_j are real since the eigenvector is, we have $\sum_{j=1}^{k} \kappa_j^2 = 1$, and so

$$\mathbf{w}^T A\mathbf{w} = \frac{\kappa_1^2}{2} \leq \frac{1}{2}$$

with equality if and only if $\kappa_1 = \pm 1$, i.e only if $\mathbf{w} = \mathbf{v}$, as required.

Thus if $\mathbf{v}$ is not an eigenvalue, all eigenvalues have modulus less than $1/2$ and so for sufficiently large r the above probability is less than $(1/2)^r$, giving the first claim of the second paragraph of the theorem.

If however $\mathbf{v}$ is an eigenvector of Q, it must have eigenvector $1/2$ as $\mathbf{v}^T Q\mathbf{v} = 1/2$ and $\mathbf{v}$ is a unit vector. Thus, as $A\mathbf{v} = 1/2\mathbf{v}$ also, we see $S\mathbf{v} = \mathbf{0}$; thus $SA\mathbf{v} = 0$ and as A kills all the other $\mathbf{e}_i$ we infer that $SA = 0$; a short calculation shows that $(SA)^T = -AS$ and so $AS = 0$ also; thus, expanding out $(A + S)^r$ by the binomial theorem, and using $AS = SA = 0$, this is

$$tr(Q^r) = tr((A + S)^r) = tr(A^r) + tr(S^r) = (1/2)^r + tr(S^r)$$

(the last equation holds as the spectrum of A is $1/2$ with multiplicity 1 and 0 with multiplicity $k - 1$). Thus the cycle is more or less likely than classically

according as $tr(S^r)$ is more or less than zero. Now S is skew so its eigenvalues are purely imaginary, and occur in conjugate pairs, which are thus negatives of each other; thus the sum of their odd powers is zero, the sum of their $(4r+2)$th powers is negative but the sum of their $4r$th powers is positive; thus cycles of length divisible by 4 are more likely than classically but those of length congruent to 2 mod 4 less likely. •

As an explicit example for the theorem, if

$$P = \begin{vmatrix} \frac{1}{2} & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{vmatrix}$$

and $\mathbf{s} = (1/4, 1/4, 1/4, 1/4)^T$ so that $Q = P/4$, it is easy to check that the eigenvalues of Q are $1/2$, 0, $i/4$ and $-i/4$ where i is a square root of -1; thus the sum of their fourth powers is easily checked to be $9/128$ which is greater than the classical probability $1/16$. (Compare the congruence condition mod 4 for being more or less probable than classically in Theorem 8.1 with the one mod 2 in Theorem 2.6; what classes of combinatorial structures give congruences mod other values r?)

We will see shortly that in the case with two equiprobable colours, these problems do not arise; cycles are always no more likely than classically, with equality if and only if we are in the classical situation.

Note also that if $\mathbf{v}$ is not an eigenvector, there can exist short cycles which are more likely than classically, so that the restriction in our theorem that the cycles be sufficiently long is genuinely necessary. Indeed if

$$P = \begin{vmatrix} \frac{1}{2} & \frac{3}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & 1 \\ 1 & 0 & \frac{1}{2} \end{vmatrix}$$

and $\mathbf{s} = (1/3, 1/3, 1/3)^T$, it is easy to check, using Theorem 8.1, that the probability of a cycle of length four is $0.0657 > 0.0625 = (1/2)^4$. However, we did not find an example where, with the top eigenvalue less than $1/2$, a cycle of odd length is more likely than classically; however, we do not see how to prove or disprove the possibility of this.

Note that here, unlike in Chapter 2, as Q is not symmetric it need not be conjugate to a diagonal matrix or have real eigenvalues. We have already illustrated the second point; for the first, in T_p with $p = 0$ it is easy to

check the eigenvalues of Q are $1/2$ with multiplicity 2, but of course Q is not conjugate to $I/2$ where I is the identity matrix.

Corollary 8.2 *The rate of decay of the probability of an r-cycle C is given, when Q is primitive, by the top eigenvalue λ of Q, in the sense that*

$$\lim_{r\to\infty} \frac{\log P\{C\}}{r} = \log(\lambda).$$

Proof. As Q is primitive, by Perron-Frobenius theory (Theorem 2.9) Q has a positive eigenvalue whose modulus is greater than that of any other eigenvalue, and the result follows as in Theorem 2.18. (No complication arises from possible complex eigenvalues; the sum of powers of eigenvalues is still real, since the complex roots occur in conjugate pairs). •

If Q is imprimitive this argument fails, and T_p with $p = 0$ shows that Q certainly can be imprimitive; however, in T_p with $p = 0$ the result still holds, as is easily checked. Finding an example where the result fails seems to be a little more difficult.

We now turn to the more explicit argument for $T_{p,q}$ (which is easily checked to give the same answer as Theorem 8.1). We will use a very easy lemma; note however there seems to be no simple analogue in ${}_sT_p$.

Lemma 8.3 *If $X_i : 1 \le i \le n$ are independent Bin$\left(\frac{1}{2}\right)$ random variables, and $S_n = \mid \{i : 2 \le i \le n, X_i \neq X_{i-1}\} \mid$ then*

$$P\{S_n = r\} = \frac{\binom{n-1}{r}}{2^{n-1}}. \bullet$$

Theorem 8.4 *Let $u = \sqrt{4pq}$. Then, in $T_{p,q}$, the probability of an r-cycle is given by*

$$\frac{1}{2}\left(\left(\frac{u+1}{4}\right)^{r-1} + \left(\frac{1-u}{4}\right)^{r-1}\right) + \frac{p+q}{2u}\left(\left(\frac{u+1}{4}\right)^{r-1} - \left(\frac{1-u}{4}\right)^{r-1}\right).$$

In addition

$$P\{1 \to 2 \to 3... \to r \mid c(1) = c(r)\} = \left(\frac{u+1}{4}\right)^{r-1} + \left(\frac{1-u}{4}\right)^{r-1}$$

$$P\{1 \to 2 \to 3... \to r \mid c(1) \neq c(r)\} = \frac{p+q}{u}\left(\left(\frac{u+1}{4}\right)^{r-1} - \left(\frac{1-u}{4}\right)^{r-1}\right).$$

Proof. Say that there is a switch between two vertices i and $i+1$ with $1 \leq i \leq n-1$ if the colours of i and $i+1$ differ. Then we have

$$P\{1 \to 2 \to 3... \to r\} = \sum_{i=0}^{r-1} P\{1 \to 2 \to 3... \to r \mid i \text{ switches}\} \frac{\binom{r-1}{i}}{2^{r-1}}$$

by Lemma 8.3 and conditioning. We need to distinguish between odd and even i. If $i = 2k$ $\left(0 \leq k \leq \frac{r-1}{2}\right)$ the pattern of transitions gets back to the colour it started at; hence, as there are $r-1-2k$ non-switches this case contributes

$$\sum_{i=0}^{\lfloor \frac{r-1}{2} \rfloor} p^k q^k \left(\frac{1}{2}\right)^{r-1-2k} \binom{r-1}{2k} \frac{1}{2^{r-1}}.$$

If $i = 2k+1$ we finish with the opposite colour; RBRB.....B and BRBR...R are equiprobable by $s = \frac{1}{2}$, the first and second cases contributing

$$\frac{p^k q^{k+1}}{2} \text{ and } \frac{p^{k+1} q^k}{2}$$

respectively; hence together these terms contribute

$$(p+q) \sum_{i=0}^{\lfloor \frac{r-2}{2} \rfloor} p^k q^k \left(\frac{1}{2}\right)^{r-2-2k} \binom{r-1}{2k+1} \frac{1}{2^{r-1}}.$$

$$\Rightarrow P\{1 \to 2 \to 3... \to r\} = \sum_{i=0}^{\lfloor \frac{r-1}{2} \rfloor} p^k q^k \left(\frac{1}{2}\right)^{r-1-2k} \binom{r-1}{2k} \frac{1}{2^{r-1}}$$

$$+\frac{1}{2} \sum_{i=0}^{\lfloor \frac{r-2}{2} \rfloor} p^{k+1} q^k \left(\frac{1}{2}\right)^{r-2-2k} \binom{r-1}{2k+1} \frac{1}{2^{r-1}}$$

$$+\frac{1}{2} \sum_{i=0}^{\lfloor \frac{r-2}{2} \rfloor} p^k q^{k+1} \left(\frac{1}{2}\right)^{r-2-2k} \binom{r-1}{2k+1} \frac{1}{2^{r-1}}$$

$$= \sum_{i=0}^{\lfloor \frac{r-1}{2} \rfloor} (4pq)^k \binom{r-1}{2k} \frac{1}{4^{r-1}} + \left(\frac{\sqrt{p}}{2\sqrt{q}} + \frac{\sqrt{q}}{2\sqrt{p}}\right) \sum_{i=0}^{\lfloor \frac{r-2}{2} \rfloor} \sqrt{4pq}^{2k+1} \binom{r-1}{2k+1} \frac{1}{4^{r-1}}$$

which, using the more convenient notation of u, is

$$= \sum_{i=0}^{\lfloor \frac{r-1}{2} \rfloor} u^{2k} \binom{r-1}{2k} \frac{1}{4^{r-1}} + \frac{p+q}{u} \sum_{i=0}^{\lfloor \frac{r-2}{2} \rfloor} u^{2k+1} \binom{r-1}{2k+1} \frac{1}{4^{r-1}}$$

since it is easy to check that

$$\frac{\sqrt{p}}{2\sqrt{q}} + \frac{\sqrt{q}}{2\sqrt{p}} = \frac{p+q}{u}.$$

As

$$\sum_{l=0}^{\lfloor \frac{m}{2} \rfloor} p^{2l} q^{m-2l} \binom{m}{2l} = \frac{(p+q)^m + (q-p)^m}{2}$$

and

$$\sum_{l=0}^{\lfloor \frac{m-1}{2} \rfloor} p^{2l+1} q^{m-2l-1} \binom{m}{2l+1} = \frac{(p+q)^m - (q-p)^m}{2}$$

the above expression is equal to

$$\frac{1}{2}\frac{(u+1)^m + (1-u)^m}{4^m} + \frac{1}{2}\frac{(u+1)^m - (1-u)^m}{4^m}.$$

The first claim is now clear, and the other two claims are immediate on noting that we have kept the terms arising from even and odd numbers of switches separate throughout the proof. •

Corollary 8.5 *The probability of a cycle in T_p is*

$$P\{1 \to 2 \to .. \to r \to 1\} = \left(\frac{u+1}{4}\right)^r + \left(\frac{1-u}{4}\right)^r$$

Proof. We note that

$$P\{1 \to 2 \to 3... \to r \to 1\} = P\{1 \to 2 \to 3... \to (r+1) \mid c(1) = c(r+1)\}$$

and the result follows from Theorem 8.4. •

Note the formula agrees with what we know in the case $u = 1$ and also when $u = 0$ so that the cycle arises only if there is only one colour, with probability $(1/2)^{r-1}$ and then all the edges go the right way; there are two possible orientations, each of which arises with probability $(1/2)^r$.

Corollary 8.6

$$P_p\{1 \to 2... \to r \to 1\}$$

is maximised, for p ranging over $[0,1]$, at $p=\frac{1}{2}$. Hence the expected number of r-cycles is maximised, for all $3 \leq r \leq n$, at $p = \frac{1}{2}$ also.

Proof. We use the fact that, if a and b are non-negative, then $(a+b)^r \geq a^r + b^r$, with equality if and only if at least one of a and b is 0. Applying this to the formula for the probability of a cycle in the previous corollary, with $a = (1+u)/4$ and $b = (1-u)/4$ we see that the probability of a cycle is at most $1/2^r$ with equality if and only if $u = 1$. The last sentence is an immediate consequence, by linearity of expectation. •

It is interesting to compare Theorem 8.4 with Theorem 2.6, giving the probability of an undirected cycle in $G_{p,q}$; there the error term was simply added on to the classical value, but here it is of a different form. In view of the link between existence of cycles and irreducibility noted above, and the obvious fact that T_p is more likely to be reducible than the classical model, Corollary 8.6 should not be surprising.

8.3 Joint probabilities of cycles

Again one ultimately wants to understand the distribution of numbers of cycles, and for this needs to understand joint probabilities of cycles. It is natural to try to apply the techniques in Theorem 3.1 to compare joint probabilities with products of individual probabilities here. The same basic idea will give some insight; however, the result we prove here will be less useful than Theorem 3.1.

We deal first with the simpler case of edge-disjoint cycles. As in Chapter 3, we use for the rest of this section the notation $P\{C\}$ to denote the probability, in whatever model of tournaments we are considering, that the tournament contains all the edges of C, that is to say the edges in question are all oriented in the way implied by the description of C.

Theorem 8.7 *Let C_1 and C_2 be two cycles which have no edge in common. Then in $T_{p,q}$, we have*

$$P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}.$$

Proof. As in the proof of Theorem 2.1, let S_i be the number of edges of C_i which are non-switches, i.e whose two vertices are the same colour, and let n_i be the number of edges in C_i, so that there are $n_i - S_i$ switches in C_i. Because the numbers of red-blue edges and blue-red edges in a cycle must be equal, there are $(n_i - S_i)/2$ red-blue edges, which arise with probability p, and the same number of blue-red edges which arise with probability q. Thus,

using the fact that C_1 and C_2 are edge-disjoint we have

$$P\{C_1 \cap C_2\} = \mathbf{E}\left(\left(\frac{1}{2}\right)^{S_1+S_2} \sqrt{pq}^{n_1+n_2-S_1-S_2}\right)$$

Since $4pq \leq 1$, we can rewrite this as

$$\sqrt{pq}^{n_1+n_2}\mathbf{E}\left(e^{\theta(S_1+S_2)}\right)$$

where

$$\theta = \log\left(\frac{1}{\sqrt{4pq}}\right) > 0.$$

(This is where we use the fact that we have insisted that the same-same edges should all still arise. In fact we could make do with the weaker condition that the probability that same-same edges arise is greater than $\sqrt{pq}$.) Also

$$P\{C_i\} = \sqrt{pq}^{n_i}\mathbf{E}\left(\theta^{S_1}\right).$$

Again as in Theorem 3.1 we set up a copy C_2^* of C_2 with neither vertices nor edges in common with C_1 and let S_2^* be the number of edges in C_2^* which are non-switches. Then, as in Theorem 3.1, considering the formula for $P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$ and using the fact that S_1 and S_2^* are independent, we need only show that

$$\mathbf{E}e^{\theta(S_1+S_2)} > \mathbf{E}e^{\theta S_1}\mathbf{E}e^{\theta S_2}.$$

Now we are in exactly the same situation as we were in the proof of Theorem 3.1 since only the colouring and the number of non-switches matter. Hence the result follows by exactly the same argument as in Theorem 3.1. •

A more important respect in which the situation differs from that in Theorem 2.1 is that the situation for cycles with edges in common is more delicate here; for clearly

$$P\{(1 \to 2 \to 3 \to 1) \cap (4 \to 3 \to 2 \to 4)\} = 0.$$

This suggests the following fix.

Definition 8.2 *In a tournament, two subgraphs C_1 and C_2 are* **consistent** *if all common edges are oriented the same way in both subgraphs.*

Then the proof above runs through for consistent cycles, using the same argument as in the last paragraph of the proof of Theorem 3.1. We record this formally;

Theorem 8.8 *Let C_1 and C_2 be two cycles in $T_{p,q}$ which are consistent. Then*

$$P\{C_1 \cap C_2\} \geq P\{C_1\}P\{C_2\}. \bullet$$

Again, as in the discussion in section 3.4 we may ask whether if consistent C_1 and C_2 have t edges in common, there exists some constant $\kappa \leq 1$ such that

$$\kappa^t P_p\{C_1 \cap C_2\} \geq P_p\{C_1\}P_p\{C_2\}.$$

In this case, an argument similar to that in Theorem 3.9 will show that we can take $\kappa = \max\{p, 1-p\}$. Again we may speculate as to whether we could sharpen this to $\kappa = 1/2$; this sharpening holds for the simplest case, of two triangles with one common edge, but we do not see how to tackle the general question.

In Corollary 3.4 we noted that

$$\mathbf{E}_{p,q}(N^r) \geq \mathbf{E}_{\alpha}(N^r) \text{ for } p > q.$$

when N is the number of cycles or k-cycles for some given k. We cannot make the analogous claim for directed cycles or directed cycles of given length here; by Theorem 8.5 no such inequality holds for the first moment, and the consistency requirement makes it hard to argue about joint probabilities and so understand higher moments.

Perhaps the most important difference between this argument and that in Theorem 3.1 is that the argument there applied to any subgraphs C_1 and C_2, but the argument here applies only to cycles, in order to ensure that we have equal numbers of red-blue and blue-red edges. This is not only a limitation of the method of proof, as the following example in T_p shows. We let the first subgraph of the tournament, C_1, consist of the two edges $1 \to 2$ and $2 \to 3$, and the second subgraph C_2 be $3 \to 1$. (The motivation for choosing this example is that one might expect that the fact that $1 \to 2$ and $2 \to 3$ are present will tend to mean that, if red-blue edges go from red to blue with probability $p > 1/2$, then 1 is likelier to be red than blue, and so that the chances of $3 \to 1$ existing are less than they would be without prior

information). Then, considering the eight possible colourings of the vertices, we see that

$$P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\}$$
$$= \frac{1/4 + 3p(1-p)}{8} - \frac{(1/2 + p + q + 2pq)(2 + 2p + 2q)}{64}$$

which is -0.060025 when $p = 0.99$ and $q = 0.01$.

In fact we can also consider the subgraphs C_1 consisting of two edges $1 \to 2$ and C_2 consisting of $2 \to 3$, when

$$P\{C_1 \cap C_2\} - P\{C_1\}P\{C_2\} = -\left(\frac{1-2p}{4}\right)^2$$

which of course is < 0 provided $p \neq 1/2$. These two examples together suggest that there is no easy generalisation of the result for cycles to more general subgraphs.

8.4 Numbers of 3-cycles.

Theorem 8.9 *Z, the number of (directed) 3-cycles in T_p has*

$$\mathbf{E}(Z) = \binom{n}{3}\frac{3u^2+1}{16} \text{ and}$$

$$Var(Z) = \frac{(n-2)(n-1)n(-26u^2n - 7u^4n + 33n + 18u^4 - 94 + 92u^2)}{512}.$$

Proof. By Theorem 8.5 and linearity of expectation we know that

$$\mathbf{E}(Z) = 2\binom{n}{3}\left(\left(\frac{u+1}{4}\right)^3 + \left(\frac{1-u}{4}\right)^3\right) = \binom{n}{3}\frac{3u^2+1}{16}$$

since there are $\binom{n}{3}$ 3-tuples and two possible orientations, namely $1 \to 2 \to 3 \to 1$ and $1 \to 2 \to 3 \to 1$.

For the variance, we need $\mathbf{E}(Z^2)$. Now Z is a sum over indicator variables of the possible cycles. If two 3-tuples have no vertices in common, or have only one vertex in common, the existence of the two cycles is independent. If they are identical, as $\binom{n}{3}$ choices are, the joint probability is 0 if the two orientations are inconsistent and is the probability of one of them if the orientations are the same. The only other case is when they have one

edge in common; then, of the four possible pairs of cycles (each triangle being orientable two ways) only two are consistent and for each consistent pair the joint probability is (using Theorem 8.4 and conditioning to get the expression)

$$\frac{1}{4}\left(\frac{1/4+pq}{2}\right)^2+\frac{1}{4}(p(\frac{p}{2})^2+q(\frac{q}{2})^2)=\frac{17-10u^2+u^4}{256}$$

so, combining the terms by computer simplification, $\mathbf{E}(Z^2)$ is

$$\frac{(n-2)(n-1)n(n^3(1+6u^2+9u^4)-n^2(27u^4+18u^2+3))}{9216}$$

$$+\frac{(n-2)(n-1)n(n(596-456u^2-108u^4)+324u^4+1656u^2-1692)}{9216}$$

$$\Rightarrow \mathrm{Var}(Z)=\mathbf{E}\left(Z^2\right)-(\mathbf{E}Z)^2$$

$$=\frac{(n-2)(n-1)n(-26u^2n-7u^4n+33n+18u^4-94+92u^2)}{512}.\bullet$$

For $u=1$ this is $n(n-1)(n-2)/32$ agreeing with [M] Theorem 10.

Corollary 8.10 $\mathbf{E}(Z)$ *is always less than classically. If* $u=1$ *Var*(Z) *takes its classical value; if* $n=3$ *and* $u<1$ *it is smaller than classically; else it is greater than classically.*

Proof. The assertion about the expectation is immediate from 8.6 and linearity of expectation. By the above calculations, the variance is greater than classically if and only if

$$\frac{(n-2)(n-1)n(-26u^2n-7u^4n+33n+18u^4-94+92u^2)}{512}\geq\frac{n(n-1)(n-2)}{32}$$

$$\Leftrightarrow n(33-26u^2-7u^4)+18u^4+92u^2-110)\geq 0.$$

$$\Leftrightarrow (1-u^2)(n(33+7u^2)-(18u^2+110))\geq 0.$$

and since $(1-u^2)>0$ for $u\neq 1$, this will then be positive for $n>4$ as is easily seen. If $n=3$ the expression is $(1-u^2)(3u^2-11)\leq 0$ with equality only if $u=1$; thus in this case the variance is always smaller than classically

(unless $u = 1$); this case could of course be deduced from the fact that the function $x \to x(1-x)$ is increasing for $x \in (0, 1/2)$. •

For $u = 0$, for example, the variance is $n(n-1)(n-2)(33n-94)/512$ which in particular is a good deal larger than classically for large n, though the method of moments estimate $P\{X = 0\} \leq \sigma^2/(\sigma^2 + \mu^2)$ still shows that the probability of no such cycles tends to 0 as $n \to \infty$.

The fact that there is no general inequality to the effect that joint probabilities exceed the product of probabilities here makes it very unlikely that the FKG inequalities or related machinery have any insight to offer here.

8.5 Estimates of the probability of irreducibility

As noted above, our models will often be more likely to be reducible than classically. It is natural to try to estimate that probability in our models. As in Chapter 6, the probability that the tournament is irreducible is only affected when the probability that red-blue edges go the one way rather than the other is very large (or very small);

Theorem 8.11 *Suppose p or $1-p$ is c/n^2, c constant. Then*

$$\lim_{n\to\infty} P\{T \text{ is irreducible}\} = 1 - e^{-c/4}$$

Proof. This is very similar to Theorem 6.10. We first note that without loss of generality $p = c/n^2$. Next, observe that with probability tending to one as n goes to infinity, there are $n/2 + o(n)$ reds and blues; in particular, with probability tending to 1, the numbers of reds and blues go to infinity with n. Now we know that a.e. tournament in the classical model is irreducible ([M, Theorem 5]). Thus, with probability tending to 1, the reds are irreducible and the blues are irreducible. Thus, with probability tending to 1, the tournament as a whole being irreducible is equivalent to not all the red-blue edges being in the one direction, as then we can indeed get from any vertex to any other. This last probability is, by the assumption,

$$1 - \sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} \left(\left(1 - \frac{c}{n^2}\right)^{i(n-i)} + \left(\frac{c}{n^2}\right)^{i(n-i)} \right).$$

Clearly only the sum

$$\sum_{i=0}^{n} \frac{\binom{n}{i}}{2^n} \left(1 - \frac{c}{n^2}\right)^{i(n-i)}$$

will contribute in the limit, by our assumption on p. But we saw in Theorem 6.9 that the value of this is $e^{-c/4}$, and the result follows. •

The possibility, implied by Theorem 8.1, that sometimes many cycles can be more likely than classically means that we have little intuition about what will happen in more general models.

8.6 The degree sequence in tournaments

Here we study the outdegrees X_i of vertex i, that is the number of vertices $j \neq i$ where the edge between i and j is oriented from i to j. We start by obtaining the probability distribution and the probability generating function of the outdegree of a single vertex.

Theorem 8.12 *In T_p we have that $P\{X_i = k\}$ is given by*

$$\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\frac{\binom{n-1}{j}}{2^{n-1}}\left(\frac{p^{k-l}(1-p)^{n-1-j-k+l}}{2^j}+\frac{(1-p)^l p^{j-l}}{2^{n-1-j}}\right)$$

and the probability generating function of X_i is

$$\frac{1}{2}\left(\left(\left(\frac{1}{4}+\frac{p}{2}\right)t+\frac{3}{4}-\frac{p}{2}\right)^{n-1}+\left(\left(\frac{3}{4}-\frac{p}{2}\right)t+\frac{1}{4}+\frac{p}{2}\right)^{n-1}\right).$$

Proof. We have

$$P\{X_i=k\}=\frac{P\{X_i=k \mid i \text{ is red}\}}{2}+\frac{P\{X_i=k \mid i \text{ is blue}\}}{2}$$

$$=\sum_{j=0}^{n-1}P\{X_i=k \mid i \text{ is red}, j \text{ reds in } \{1,2,\ldots n\}\backslash i\}\frac{\binom{n-1}{j}}{2^n}$$

$$+\sum_{j=0}^{n-1}P\{X_i=k \mid i \text{ is blue}, j \text{ reds in } \{1,2,\ldots n\}\backslash i\}\frac{\binom{n-1}{j}}{2^n}$$

$$=\sum_{j=0}^{n-1}\sum_{l=0}^{k}P\{\text{Bin}(j,\frac{1}{2})=l \text{ and } \text{Bin}\,(n-1-j,p)=k-l\}\frac{\binom{n-1}{j}}{2^n}$$

$$+\sum_{j=0}^{n-1}\sum_{l=0}^{k}P\{\text{Bin}\,(j,1-p)=l \text{ and } \text{Bin}(n-1-j,\frac{1}{2})=k-l\}\frac{\binom{n-1}{j}}{2^n}.$$

The binomials involved are independent, so this is

$$\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\frac{\binom{n-1}{j}}{2^{n-1}}\left(\frac{p^{k-l}(1-p)^{n-1-j-k+l}}{2^j}+\frac{(1-p)^l p^{j-l}}{2^{(n-1-j)}}\right).$$

For the second statement, the probability generating function is

$$\sum_{k=0}^{n-1}P\{X_i=k\}t^k$$

$$=\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\binom{n-1}{j}p^{k-l}(1-p)^{n-1-j-k+l}\left(\frac{1}{2}\right)^j t^k$$

$$+\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\binom{n-1}{j}p^{j-l}(1-p)^l\left(\frac{1}{2}\right)^{n-1-j}t^k$$

$$=\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\binom{n-1}{j}(pt)^{k-l}(1-p)^{n-1-j-(k-l)}\left(\frac{1}{2}\right)^j t^l$$

$$+\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=0}^{k}\binom{j}{l}\binom{n-1-j}{k-l}\binom{n-1}{j}p^{j-l}(t(1-p))^l\left(\frac{1}{2}\right)^{n-1-j}t^{k-l}$$

which, shifting the variable in the innermost summation of the top line from l to $k-l$ is

$$=\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\binom{j}{l}(pt+1-p)^{n-1-j}\binom{n-1}{j}\left(\frac{1}{2}\right)^j t^l$$

$$+\frac{1}{2^n}\sum_{k=0}^{n-1}\sum_{j=0}^{n-1}\binom{n-1-j}{k-l}(p+t(1-p))^j\left(\frac{1}{2}\right)^{n-1-j}t^{k-l}$$

which is, summing out the l variable,

$$=\frac{1}{2^n}\sum_{j=0}^{n-1}(pt+1-p)^{n-1-j}\binom{n-1}{j}\left(\frac{1+t}{2}\right)^j$$

$$+\frac{1}{2^n}\sum_{j=0}^{n-1}\binom{n-1}{j}(p+t(1-p))^j\left(\frac{1}{2}\right)^{n-1-j}(1+t)^{n-1-j}$$

which, simplifying further and tidying up is

$$\frac{1}{2^n}\left(\left(pt+1-p+\frac{1+t}{2}\right)^{n-1}+\left((1-p)\,t+p+\frac{1+t}{2}\right)^{n-1}\right)$$

$$=\frac{1}{2}\left(\left(\left(\frac{p}{2}+\frac{1}{4}\right)t+\frac{3}{4}-\frac{p}{2}\right)^{n-1}+\left(\left(\frac{3}{4}-\frac{p}{2}\right)t+\frac{p}{2}+\frac{1}{4}\right)^{n-1}\right)$$

as required. •

Since for each pair i,j $P\{i \to j\} = 1/2$, we have that $\mathbf{E}(X_i) = (n-1)/2$; it is easy to check that this is also the first derivative of the above generating function evaluated at 1. Similarily, the variance can be obtained from the probability generating function; it is

$$\frac{(n-1)(1+4p(1-p))}{8}+\frac{n(n-1)(1-4p(1-p))}{16}.$$

Thus the variance grows like n^2 rather than n as soon as p ceases to be exactly $1/2$; this is as one would expect, since there will then be a bimodal distribution of the outdegrees, with the reds tending to have outdegrees near to $(n-1)(p/2+1/4)$ and the blues tending to have outdegrees near to $(n-1)(3/4-p/2)$, so most vertices will have degree away from the mean degree by about $\mid (1/2-p)n \mid$, rather than about $\sqrt{n}$ in the classical situation.

Knowledge of the moment generating function also allows us to understand the probability of a large deviation in the degree of a vertex. This is again an application of the Gartner-Ellis theorem, although the details here are somewhat easier than in Chapter 4. We shall only work it out in the case when $p \geq 1/2$; this is no real loss of information as we have noted already, and it will save having to write out some cumbersome formulae which could in any case be derived by exactly the same techniques as we use.

Lemma 8.13 *In T_p, $p > 1/2$, letting ϕ_n be the cumulant generating function of the outdegree of a vertex, and $\phi = \lim_{n\to\infty} \phi_n/n$,*

$$\phi(\theta) = \log\left(\left(\frac{p}{2}+\frac{1}{4}\right)e^{\theta}+\frac{3}{4}-\frac{p}{2}\right) \text{ for } \theta > 0,$$

$$\phi(\theta) = \log\left(\left(\frac{3}{4}-\frac{p}{2}\right)e^{\theta}+\frac{p}{2}+\frac{1}{4}\right) \text{ if } \theta < 0$$

and is zero if $\theta = 0$.

Proof. We first assume that $\theta > 0$, so that $e^\theta > 1$ and so it is easy to check that

$$\left(\frac{p}{2}+\frac{1}{4}\right)e^\theta + \frac{3}{4} - \frac{p}{2} \geq \left(\frac{3}{4}-\frac{p}{2}\right)e^\theta + \frac{p}{2} + \frac{1}{4} \geq 0.$$

since $p > 1/2$. In our previous notation, we have

$$\phi = \lim_{n\to\infty} \frac{\phi_n}{n}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{2}\left(\left(\left(\frac{p}{2}+\frac{1}{4}\right)e^\theta + \frac{3}{4} - \frac{p}{2}\right)^{n-1} + \left(\left(\frac{3}{4}-\frac{p}{2}\right)e^\theta + \frac{p}{2} + \frac{1}{4}\right)^{n-1}\right)\right)$$

$$= \lim_{n\to\infty} \frac{-\log 2}{n} + \frac{n-1}{n} \log\left(\frac{1}{2}\left(\left(\frac{p}{2}+\frac{1}{4}\right)e^\theta + \frac{3}{4} - \frac{p}{2}\right)(1+x_n)\right)$$

where $0 < x_n < 1$ by the assumption, and $x_n \to 0$ as $n \to \infty$. Thus, taking the limit, we see that

$$\phi(\theta) = \log\left(\left(\frac{p}{2}+\frac{1}{4}\right)e^\theta + \frac{3}{4} - \frac{p}{2}\right)$$

The other case is when $\theta < 0$ so that

$$0 \leq \left(\frac{p}{2}+\frac{1}{4}\right)e^\theta + \frac{3}{4} - \frac{p}{2} \leq \left(\frac{3}{4}-\frac{p}{2}\right)e^\theta + \frac{p}{2} + \frac{1}{4}$$

(again by $p > 1/2$); then the obvious analogous argument yields

$$\phi(\theta) = \log\left(\left(\frac{3}{4}-\frac{p}{2}\right)e^\theta + \frac{p}{2} + \frac{1}{4}\right). \bullet$$

Again we must investigate the differentiability properties of ϕ.

Lemma 8.14 *ϕ is differentiable away from the origin, and the derivative takes every value in $[0, \frac{3}{4}-\frac{p}{2}]$ and $[\frac{1}{4}+\frac{p}{2}, 1]$.*

Proof. Only the assertion about non-differentiability at $\theta = 0$ requires comment. We have

$$\lim_{\theta\to 0+} \frac{\phi(\theta)-\phi(0)}{\theta}$$

$$= \lim_{\theta\to 0} \frac{\log\left(\left(\frac{3}{4}-\frac{p}{2}\right)e^\theta + \frac{p}{2} + \frac{1}{4}\right)}{\theta}$$

$$= \lim_{\theta \to 0} \frac{\log\left(\left(\frac{3}{4} - \frac{p}{2}\right)\left(e^{\theta} - 1\right) + 1\right)}{\theta}$$

which, using the expansions of the exponential function and $\log(1+x)$ is

$$= \lim_{\theta \to 0} \frac{\left(\frac{3}{4} - \frac{p}{2}\right)(\theta + \theta^2 + ...) + \left(\left(\frac{3}{4} - \frac{p}{2}\right)(\theta + \theta^2 + ...)\right)^2 + ...}{\theta} = \frac{3}{4} - \frac{p}{2}$$

and an identical argument shows that the left derivative is $1/4 + p/2$. The result follows. •

So, as in Chapter 5, unless $p = 1/2$, there is some interval where the rate function gives no information about what is happening. We would again guess that in that region, the rate function is obtained by dividing the probability of the large deviation just by the constant 1. For, noting that $(p/2 + 1/4) > a > (3/4 - p/2)$, we see that if the vertex is red (which happens with probability $1/2$), its degree will be about $(1/4 + p/2)(n-1)$ with probability tending to 1, so with probability tending to 1 a red vertex has degree greater than an; on the other hand, a similar argument shows that with probability tending to 1, a blue vertex has degree less than an. We summarise all this;

Theorem 8.15 *Let X_i be the outdegree of a vertex in T_p, where without loss of generality $p > 1/2$. Then, for $a \in [0, (3/4 - p/2)]$ or $[(1/2 + p/4), 1]$, we have*

$$\lim_{n \to \infty} \frac{\log\left(P\{X_i \geq an\}\right)}{n}$$

is the maximum of

$$\sup_{\theta > 0} \left(\theta y - \log\left(\left(\frac{3}{4} - \frac{p}{2}\right)e^{\theta} + \frac{p}{2} + \frac{1}{4}\right)\right), 0 \text{ and}$$

$$\sup_{\theta < 0} \left(\theta y - \log\left(\left(\frac{3}{4} - \frac{p}{2}\right)e^{\theta} + \frac{p}{2} + \frac{1}{4}\right)\right).$$

For other values of a,

$$\lim_{n \to \infty} P\{X_i \geq an\} = \frac{1}{2}.$$

Proof. This is clear from the Gartner-Ellis theorem (Theorem 5.3) and the foregoing remarks. •

It is desirable to understand the rate function somewhat more explicitly. By elementary calculus, the function

$$f(\theta) = \theta y - \log\left(\left(\frac{3}{4} - \frac{p}{2}\right)e^{\theta} + \frac{p}{2} + \frac{1}{4}\right)$$

has its unique turning point at

$$\theta = \log\left(\frac{y(1+2p)}{(1-y)(3-2p)}\right)$$

and this is a maximum; the value at which the turning point occurs is positive if and only if $y > (3-2p)/4$; if that holds, we get

$$\sup_{\theta>0}\left(\theta y - \log\left(\left(\frac{3}{4} - \frac{p}{2}\right)e^{\theta} + \frac{p}{2} + \frac{1}{4}\right)\right)$$

$$= y\log(y)+y\log(1+2p)-y\log(1-y)-y\log(3-2p)+\log(4(1-y))-\log(1+2p).$$

Similarily the function

$$f(\theta) = \theta y - \log\left(\frac{1}{4} + \frac{p}{2} + e^{\theta}\left(\frac{3}{4} - \frac{p}{2}\right)\right)$$

has its unique turning point at

$$\theta = \log\left(\frac{y(3-2p)}{(1-y)(1+2p)}\right)$$

and this is a maximum too; the value at which the turning point occurs is negative if and only if $y < (1+2p)/4$, and if this holds we get

$$\sup_{\theta>0}\left(\theta y - \log\left(\left(\frac{1}{4} + \frac{p}{2}\right)e^{\theta} + \frac{3}{4} + \frac{p}{2}\right)\right)$$

$$= y\log(y)+y\log(3-2p)-y\log(1-y)-y\log(1+2p)+\log(4(1-y))-\log(3-2p).$$

As with the degrees in $G_{p,q}$, the outdegrees in T_p do not show up the correlation structure in their pairwise correlation.

Lemma 8.16

$$Corr(X_i, X_j) = \frac{-1}{n-1} \text{ in } T_p \forall p \in [0,1].$$

Proof. We have

$$\sum_{i=0}^{n} X_i = \frac{n(n-1)}{2} \text{ constant}$$

and so, taking variances on both sides and using the fact that the outdegrees are exchangeable, for any $i \neq j$ we have

$$n(n-1)\,\mathrm{Cov}(X_i, X_j) + n\mathrm{Var}(X_i) = 0$$

$$\Rightarrow \mathrm{Corr}(X_i, X_j) = \frac{-1}{n-1}$$

as required. •

It is thus again natural to ask what we can say about the maximum or minimum outdegree in our models. Classically, Theorem 29 of [M] shows that it is near to

$$(n-1)/2 + \sqrt{(n-1)\log(n-1)/2}.$$

We show how to give an upper and lower bound on this quantity in our models. We can suppose $p > 1/2$ without loss of generality; thus the maximum degree occurs, with overwhelming probability, in the red vertices. We first try to bound the maximum outdegree below. If we consider a vertex of maximum red-red degree, we get a lower bound of

$$(n/2-1)/2 + \sqrt{(n/2-1)\log(n/2-1)/2} + pn/2 + \text{ lower order terms}$$

$$= (p/2 + 1/4)n + \sqrt{n\log(n)/4} + \ldots$$

and if we consider a vertex of maximum red-blue degree, that is the maximum of $n/2$ independent $\mathrm{Bin}(n/2, p)$, we get, as in Chapter 4

$$pn/2 + \sqrt{2p(1-p)\log(n/2)n/2} + (n/2)/2 + \text{ lower order terms}$$

$$= (p/2 + 1/4)n + \sqrt{p(1-p)n\log(n)} + \ldots$$

and so (as $p(1-p) \leq 1/4$) it is the first of these which gives the better bound. The same argument shows that an upper bound is

$$(p/2 + 1/4)n + \sqrt{n\log(n)}(1/2 + \sqrt{p(1-p)}) + \ldots$$

but we are not aware of any detailed information on the rest of the outdegree sequence, and so cannot produce as precise a result as in Theorem 4.21. Note that as $1-p$ gets close to zero, the term is close to the maximum red-red degree, as one would expect; but in general, the classical case suggests that neither bound is terribly good.

8.7 A random variable related to the orientations of the edges

In Chapters 4 and 5 we considered the number of edges in undirected graphs. Because in a tournament all $n(n-1)/2$ edges are always present, the number of edges is a rather boring random variable. However we can instead consider, since our tournaments are labelled, the number Z of pairs of vertices $i < j$ for which the edge goes $i \to j$ (more generally, we could consider any $n(n-1)/2$ pre-defined directions; in our models, the answer **would** depend on the choice of directions, as we shall see below, although classically of course it would not do so). Thus Z could be seen as measuring how far some labelled tournament is from a certain model; for example, we might be comparing a student's rankings of some objects with those of an established expert. Classically Z is a sum of $n(n-1)/2$ indicators of independent Bernoulli trials so standard theory applies. In T_p the random variables are again each 1 or 0 equiprobably, but are now dependent; for example a short calculation shows that with three vertices, the probability that the edges go $1 \to 2$, $1 \to 3$ and $2 \to 3$ is $1/8 + (1/4 - p(1-p))/8$ which is always at least as large as its classical value $1/8$. If however we had instead asked for the probability that the edges go $1 \to 2$, $2 \to 3$ and $3 \to 1$, it would be, by Theorem 8.4, equal to $(12p(1-p)+1)/32$ which is **smaller** than classically. Note however that relabelling the vertices does not affect matters.

We henceforth deal only with the ordering where we ask how many edges go from i to j where $i < j$. We already know $\mathbf{E}(Z)$. What is its variance?

Lemma 8.17

$$Var(Z) \text{ in } T_p = \frac{n(n-1)}{4} \text{ independent of } p.$$

Proof. $Z = \sum_{1 \le i < j \le n} X_{ij}$, where $X_{ij} = 1$ if the edge between i and j goes $i \to j$. Hence

$$\mathrm{Var}(Z) = \mathrm{Cov}(Z, Z) = \mathrm{Cov}\left(\sum_{1 \le i < j \le n} X_{ij}, \sum_{1 \le k < l \le n} X_{kl} \right)$$

$$= \sum_{1 \le i < j \le n} \sum_{1 \le k < l \le n} \mathrm{Cov}(X_{ij}, X_{kl}).$$

Now X_{ij} and X_{kl} are independent (and so have covariance zero) unless they have one or more vertices in common. The cases to consider are

1. X_{ij} and X_{il} with $(j \neq l)$, of which there are in total

$$\sum_{i=1}^{n-2} i(n-i)(n-i-1) = \frac{(n+1)n(n-1)(n-2)}{12}$$

cases and the covariance is $(2p-1)^2/16$ in each by a short calculation.

2. X_{ij} and X_{kj} with $i \neq k$; again there are $(n+1)n(n-1)(n-2)$ cases in total by a similar argument, with covariance $(2p-1)^2/16$ in each case.

3. X_{ij} and X_{jk} with $i < j < k$, of which there are in total

$$\sum_{j=2}^{n-1} j(j-1)(n-j) = \frac{(n+1)n(n-1)(n-2)}{12}$$

cases, and here the covariance is easily checked (as in the calculations illustrating that Theorem 8.8 is limited to cycles) to be $-(2p-1)^2/16$.

4. X_{ij} and X_{ki} with $k < i < j$; again there are $(n+1)n(n-1)(n-2)/12$ cases, each with covariance $-(2p-1)^2/16$.

5. X_{ij} and X_{ij}; in each of the $n(n-1)/2$ cases we have

$$\mathrm{Cov}\,(X_{ij}, X_{ij}) = \mathbf{E}X_{ij} - \mathbf{E}\,(X_{ij})^2 = \frac{1}{4}$$

We thus see that cases 3 and 4 cancel out cases 1 and 2, leaving us with variance $n(n-1)/8$ which is also the variance of $n(n-1)/2$ independent Bernoulli trials with probability $1/2$, as required. •

This result is analogous to Theorem 4.5, though the method of proof is a little different. Again higher moments will be different from classically.

It seems likely that any study of large deviations in Z will depend heavily on how many of the first few vertices are coloured red, with a large deviation in them giving rise to a large deviation in the number of edges going from the lesser label to the greater (assuming $p > 1/2$), and as in Chapter 5 we will again get this less expensively than usual. However this situation seems less amenable to exact results than the previous one.

9 Epilogue

9.1 Summary and directions for future work

In this short chapter we briefly review what has been achieved in this thesis and indicate some possible directions for future work.

We mentioned in the introduction that there seems to be no previous literature on the subject; thus the results are new (except of course, where we quote standard facts or results of previous authors). Three main limitations of the results merit comment. First, as mentioned in the introduction, much of Chapters 2-5 is in some sense results preliminary to developing a theory of our random graphs, with the corresponding questions for the classical model being trivial or easy, and so this material has a different feel from classical random graph theory. Second, some of the results (for example, the material on the maximum degree in Chapter 4 and the result on the probability of connectedness in $G_{p,q}$ with q small in Chapter 6) rely heavily on exploiting much more detailed information about what happens in the classical model, to get a result for our model which is often substantially less precise. Thirdly, note that often we have only proved results for a limited range of values of the parameters; for example, the material in Chapter 4 on the maximum degree was only developed for p and q constant, but the classical results on which we relied work in far greater generality. Some of these extensions may be pretty easy; others likely will not.

In Chapter 2, we discussed the probabilities of trees and cycles in our models, and how they compare with the corresponding classical values. A reasonably satisfactory solution was obtained in many respects; the main remaining problem is of course to resolve whether or not trees are always at least as likely to arise as classically, and if not to understand as far as possible when they will be more likely and when less likely. If the conjecture that they are always at least as likely as classically turns out to be false, it seems too much to hope for a simple categorisation of when they are more probable than classically, but one would hope that more general partial results than we have at the moment might be possible.

In Chapter 3, we discussed when the joint probability of two subgraphs arising is greater or less than the product of their individual probabilities. There is some scope for seeing how much more generally we can get the technique of Theorem 3.1 to work; but probably more important in the long term is working out to what extent some result similar to the Janson inequalities

is applicable in our models.

In Chapter 4, perhaps the most interesting features are the work on the maximum degree in graphs. There is scope for considering how far these techniques can be generalised to deal with, say, non-constant values of the parameters p and q, or indeed more general parameters.

In Chapter 5, the main priority for future work is to see whether we can get a more detailed understanding of the large deviations theory in the general case, the $G_{p,q}$ case now being fairly fully understood. There may well be some fairly substantial problems with this.

In Chapter 6, one topic which merits investigation is understanding, for α the threshold probability for connectedness, the behaviour of the limiting probability of connectedness, and in particular whether it is constant (in $G_{p,q}$) for $q > p$. Another such is the topic of the eigenvalue distribution for adjacency matrices in our models; classically this subject is well understood; see for example [B] Theorem XIV.12 and XIV.13, but it is not clear what will happen in here.

Chapter 7 is, as we observed in it, a rather more experimental chapter than some of the others, and there is plenty scope for further work on the topics in it, including precise results on when the distribution of clique sizes is multimodal.

In Chapter 8 again there are some questions about generalisation of the results, but perhaps more important is to start moving beyond the techniques we employed, which as we said are in general primarily modifications of techniques in earlier chapters, to prove results which in some cases are more in the spirit of traditional random tournament theory.

9.2 Applications and statistical questions

We have paid scant regard to the various possible applications, but there is plenty of scope for work on these too. One such project would be to use suitable random graphs from models of our kinds to generate patterns of ESSs, in the manner implied by Theorem 7.1, to investigate whether or not certain patterns of ESSs are attainable.

Another such is the modelling of the spread of infectious diseases. Various models have been considered (see e.g [B], Chapter XIV, section 5). Another model is discussed by Barbour and Mollison [BM] who show that the classical model $G(n,p)$ is essentially equivalent to the so called Reed-Frost model, a standard elementary epidemic model consisting of a Markov chain with states

$\{(i,r) : i,r \geq 0, i+r \leq n\}$ where, for $0 \leq j \leq n-i-r$,

$$P\{(i,r) \to (j,i+r)\} = \binom{n-i-r}{j} \left(1-(1-p)^i\right)^j (1-p)^{i(n-i-r-j)}$$

where i is the number of infected individuals, r the number removed. The construction is as follows; we take one or more vertices as the initial infected individuals, numbering $i(0)$; their neighbours in the graph are the $i(1)$ infectives at time 1, and then the infectives at time 0 are removed. In general, $i(t+1)$ is the number of neighbours of the $i(t)$ infectives at time t who have not previously been infected. Note that this set-up emphasises the role of individuals, and the lists L_a of those infected by a particular individual a. This process on the surface of things corresponds to a digraph; however we note that in an epidemic, just one of the two events a infects b and b infects a occurs, so that it makes no difference to the process to make the events that $b \in L_a$ and $a \in L_b$ dependent, provided the probability of the event remains unchanged, and that the events remain independent of all other events. Thus we need only consider events that a infects b or b infects a, which still happen independently with probability p, and so we can ignore the orientations of the edges, and so we arrive at the (undirected) random graph $G(n,p)$. Similarily the number and orders of components of a graph in $G(n,p)$ can be constructed using a Reed-Frost epidemic model. The interaction between the two subjects goes both ways; for example, an old result of Daniels to the effect that the number of survivors in an epidemic is asymptotically Poisson is here seen to be an easy consequence of the result on the number of isolated vertices in $G(n,p)$, and in the other direction one can recover the asymptotic order of the giant component of $G(n,p)$ when $p = c/n$ from the so-called branching process approximation to

$$P\{\sum_{t\geq 0} i(t) > n(1-1/c) \mid i(0) = 1\}.$$

Barbour and Mollison remark that the connections developed in their paper lead naturally to the conjecture that the distance from a randomly chosen vertex in the giant component to one of the vertices furthest from it is $k\log(n) + O(1)$ with variability confined to the $O(1)$ term, a conjecture which had not been obvious from purely random-graph-theoretic considerations. They also show that other more detailed information about the order of the giant component can be obtained from more detailed study of the Reed-Frost epidemic.

In all the above models, it will be noted that we have treated all individuals as being equally likely to catch and transmit the disease. However, this assumption is patently unsustainable in practice; for a simple example, sexually transmitted diseases pass primarily from one sex to the opposite one, and this suggests that some model along our lines might yield a better approximation to what is going on here.

It is clear that $\Gamma(n, k, \mathbf{s}, P)$ corresponds to a multitype Reed-Frost model, where, before the process starts at all, we independently assign one of k types randomly to each individual, and then states of the process are given by having, for each $1 \leq l \leq k$, numbers i_l of infectives of type l and r_l of removed individuals of type l (so that the number of susceptibles of type l, which we will here denote z_l to avoid confusion with the s_l, the probabilities that a vertex is of colour l, is equal to $n_l - i_l - r_l$) and then the system changes states with probabilities given by

$$z_l(t+1) \sim \mathrm{Bin}(z_l(t), \prod_{j=1}^{k} (1-p_{lj})^{i_j(t)})$$

with

$$i_l(t+1) = z_l(t) - z_l(t+1) \text{ and } i_l(0) = m_l, s_l(0) = n_l.$$

Again of course we are looking at the situation in monochrome, so we would want to add up the i_l etc. over all values of l. A minor irritation here is that strictly speaking we cannot insist in advance that we have some number m_l of colour l infectives, since of course the colouring process may in principle not give us that many vertices of that colour. This will present no problems if the idea is to start with a certain number of infectives, and then let them take types randomly, which is closer to the spirit of only looking at the situation in monochrome, but sometimes in practice we may have prior information about what kind of individuals set the process in motion. We may well be able to get round this problem for many asymptotic arguments, if the $i_l(0)$ are fixed or only grow slowly while n goes to infinity by some kind of argument replacing the process with an approximation to it which always has at least $i_l(0)$ vertices of colour l and proving that the differences between the two processes are, in the limit, negligible. The analogous situation for given $m_l(0)$ and $n_l(0)$ has been studied by Scalia-Tomba [ST], who obtains asymptotics for the final size of the epidemic. Extensions of this work have been carried out by Andersson, and Ball and Clancy.

We have also considered only probabilistic questions. However there are natural statistical questions arising, as to how best to estimate the parameters of an RRC model given a sample of (labelled as usual) graphs which we believe to be from some such model (and which we see in monochrome). Classically this is just estimating the success probability in independent Bernoulli trials, and any amount of theory says that the obvious estimate is in any number of ways the best one; but in our case, it is easy to see that the full likelihood of getting some particular graph can only be written as a sum over all the possible colourings, which makes maximum likelihood estimation (for example) a rather daunting prospect. Naive estimators also present some difficulties; for an (over)-simple example, if we believe a sample of one graph has arisen from a $G_{p,q}$ model, and try to estimate p and q by counting the number A of edges and the number B of triangles, and then solving the equations $\binom{n}{2}(p+q)/2 = A$ and $\binom{n}{3}(p^3+3pq^2)/4 = B$, then if the sample consists of the graph on three vertices with edges 1-2 and 1-3, but not 2-3, we would get from the second equation that $p = q = 0$ contradicting the first equation. In summary, there is scope for development of sensible methods for addressing these problems.

9.3 Bibliography

[A] Aldous D. Aspects of exchangeability. In: Lecture Notes in Mathematics 1117. Springer, Berlin (1985).
[Al] Alon, N. Eigenvalues and expanders. Combinatorica 6 (1986) 83-96.
[ASE] Alon, N; Spencer, J; Erdos, P. The Probabilistic Method. John Wiley, New York, (1991).
[An] Anderson, I. Combinatorics of finite sets. Clarendon Press, Oxford (1986).
[AR] Appel, M. J. B; Russo, R. P. The maximum vertex degree of a graph on uniform points in $[0,1]^d$. Adv. Appl. Prob. 29 (1997) 567-581.
[B] Bollobas, B. Random Graphs. Academic Press (1985).
[B1] Bollobas, B. (Editor). Probabilistic Combinatorics and its Applications. Proceedings of Symposia in Applied Mathematics Volume 44. American Mathematical Society, Providence, Rhode Island, (1991).
[B2] Bollobas, B. Combinatorics. Cambridge University Press (1986).
[BB] Biggins, J. D., Bingham, N. H. Large deviations in the supercritical branching process. Adv. Appl. Prob. 25 (1993) 757-772.
[BC] Bishop, D. T; Cannings, C. Models of animal conflict. Adv. Appl. Prob 8 (1976) 616-621.
[BCV] Broom, M; Cannings, C; Vickers, G. T. On the number of local maxima of a constrained quadratic form. Proc. R. Soc. Lond. A (1993) 443, 573-584.
[BGJ] Bollobas, B.; Grimmett, G. R.; Janson, S. The random-cluster model on the complete graph. Probability Theory and Related Fields 104 (3) (1996).
[BHJ] Barbour, A. D, Holst L., Janson S. Poisson Approximation. Oxford Studies in Probability 2. Clarendon Press, Oxford, (1985).
[BM] Barbour, A. D; Mollison, D. Random graphs and epidemics. In; Lecture Notes in Biomathematics 89 (Lefevre and Gabriel, Eds.). Springer (1987)
[CO] Cremona, J. E.; Odoni, R. W. K. Some density results for negative Pell equations; an application of graph theory. Journal LMS 39 (1989) 16-28.
[CV] Cannings, C.; Vickers, G. T. Patterns of ESSs, I, II. J. Theoret. Biol. 132 (1988) 387-408 and 409-420.
[D] Durrett, R. Probability: Theory and Examples (2nd edition). Duxbury Press 1996.
[DZ] Dembo, A; Zeitouni, O. Large Deviation techniques. Jones and Bartlett, Boston, 1993.

[Fe] Feller, W. An Introduction to Probability Theory and its Applications. Volume 1: Third Edition. John Wiley, 1958.
[GM] Grimmett, G. R.; McDiarmid, C. On colouring random graphs. Math. Proc. Camb. Phil. Soc. 77 (1975) 313-324.
[GS] Grimmett, G. R., Stirzaker D. Probability and Random Processes. 2nd edition. Oxford University Press, 1994.
[Ha] Harris, T. E. A lower bound for the critical probability in a certain percolation process. Proc. Camb. Phil. Soc. 56 (1960) 13-20.
[HJ] Horn, R; Johnson, C. R. Matrix analysis. CUP (1985).
[JKLP] Janson, S; Knuth, D E; Luczak, T; Pittel, B. The birth of the giant component. Random Struct. Algorithms. 4 (1993) 233-358.
[Ju] Juhasz, F. The asymptotic behaviour of Fiedler's algebraic connectivity. Discrete Math. 96 (1991) 59-63.
[Ke] Kelmans, A. K. On the connectedness of random graphs. Automatika i Telemech. 28 (1967) 567-74. (Russian)
[Ki] Kingman, J. F. C. A matrix inequality. Quart. J. Math (Oxford) 12 (1961), 78-80.
[KLW] Klee, V. L.; Larman, D. G.; Wright, E. M. The proportion of labelled bipartite graphs which are connected. Journal LMS 24 1981, 397-404.
[Ko] Kovalenko, I. N. On the structure of random directed graphs. Theory Probab. Math. Statist. 6 (1975) 83-92.
[Le] Leader, I. B. Discrete isoperimetric inequalities. In [B1].
[McD] McDiarmid, C. On the method of bounded differences. In; Surveys in Combinatorics 1989 (J. Siemons ed.), LMS Lecture Note Series, vol. 141, CUP (1989) pp 148-188.
[M] Moon, J. W. Tournaments. Mimeographed notes. 1968.
[MS] Mulholland, M. P; Smith, C. A. B. An inequality arising in genetical theory. Amer. Math. Monthly 66 (1959), 673-683.
[O] O'Connell, N. Some large deviation results for sparse random graphs. Technical report HPL-BRIMS-96-22. To appear in Probab. Theory. and Rel. Fields (available at http://hplbwww.hpl.hp.com/brims/members/noc-own.html.)
[Pl] Plesnik, J. Critical graphs of given diameter. Acta Fac. Rev. Nat. Univ. Comen. Math 30 (1975) 71-93
[RW] Ruczinski, A.; Wormald, N. C. Random graph processes with maximum degree 2. Ann. Appl. Prob. (1997) 183-200.
[Sa] Saltyakov, A. I. The number of components in a random bipartite graph. Discrete Math. Appl. 5 (1995) 515-523.

[SS] Sasser D. W.; Slater, M. L. On the inequality

$$\sum_{i=1}^{k} x_i y_i \geq \frac{1}{n} \sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i$$

and the van der Waerden permanent conjecture. Journal of Combinatorial Theory 3 (1967), 25-33.
[Se] Seneta, E. Non-negative matrices. (1st edition). George Allen and Unwin, 1973.
[Si] Silverman, B. W. Limit theorems for dissociated random variables. Ann. Appl. Probab. 8 (1976) 806-819.
[S] Spencer, J. Modern probabilistic methods in combinatorics. In; Surveys in Combinatorics 1995 (P. Rowlinson, Ed.). LMS Lecture Note Series, 218. CUP, 1995.
[ShTa] Shmoys, D. B; Tardos, E. Computational Complexity. In: Handbook of Combinatorics (Vol 2.) (R. L. Graham, M. Grotschel and L. Lovasz, Eds.), North-Holland (1995) 1599-1646.
[ST] Scalia-Tomba, G. Asymptotic final size distribution of the multitype Reed-Frost process. J. Appl. Probab. 23 (1986) no. 3, 563-584.
[T] Thomason, A. G. A graph property not satisfying a "zero-one" law. European. J. Combin. 9 (1988) 517-522.
[Ta] Takacs, L. Queues, ballots and random graphs. J. Appl. Prob. 26 (1989) 103-112.
[VW] Van Lint, J. H. and Wilson, R. M. A course in combinatorics. CUP (1992).
[W] Woodcock, C. F. On the asymptotic behaviour of a function arising from tossing coins. Bull. LMS Volume 28 part 1 (1996) 19-24.
[We] Webster, R. J. Convexity. OUP (1994).