Cliques decomposition of random graphs with exchangeable hidden colors\(^*\)\(^†\)

**Annalisa Cerquetti\(^†\)**

*Istituto di Metodi Quantitativi, Università L. Bocconi, Milano, Italy*

**Abstract**

Random graphs with exchangeable hidden colors have been recently introduced in Cerquetti and Fortini (2003). Here we restrict to the subclass driven by a complete graph. Due to their intrinsic structure, the elements of such class decompose naturally into cliques, (maximal complete subgraphs). Relying on Kingman’s theory of exchangeable random partitions, we derive some distributional results for cliques decomposition under the assumption that vertices colors are a sample from a Dirichlet process. We show that the distribution of the normalized cliques orders, in order of their least vertex, converges to a GEM distribution, and that the law of the ranked normalized cliques orders has a Poisson-Dirichlet limit as \(n\), the number of vertices, tends to infinity. We also prove that, for every \(n\), the clique counts vector has joint distribution the Ewens Sampling Formula, hence the order of magnitude of the expected number of cliques in the decomposition is asymptotically \(\log n\). Thus such class provides an example of logarithmic combinatorial random graphs.

**1 Introduction**

In general every probability space whose points are graphs, gives us the notion of a random graph. Given a finite set of vertices \([n] = \{1, \ldots, n\}\), the classical (finite, undirected, acyclic, purely) random graph (Erdős and Rényi, 1960) is defined by the space \(\mathcal{G}(n, p)\) for \(0 \leq p \leq 1\), whose random elements are obtained by selecting each edge independently, with probability \(p\). One of the most appealing generalization of Erdős standard model is obtained by randomly coloring vertices according to a color distribution and deleting edges independently with color dependent probabilities. (See e.g. Penman, 1998; Söderberger, 2003). In Cerquetti and Fortini (2003) a different generalization is provided assuming vertices colors are exchangeable.

\(^*\)Research partially supported by MIUR 2002, Research program: *Bayesian nonparametric methods and their applications*.


\(^‡\)Corresponding author: Istituto di Metodi Quantitativi, Viale Isonzo, 25, 20133 Milano, Italy. E-mail: annalisa.cerquetti@unibocconi.it
Here we focus on a particular subclass of the family of random graphs with exchangeable hidden colors (ECRG), the one arising when the underlying fixed graph is totally connected. By construction, each element of this class can be easily interpreted as a graphical representation of an exchangeable random partition (Kingman, 1978; Aldous, 1985). In fact, edges arise among vertices whose colors, which are a sample from a sequence of exchangeable random variables, coincide, hence connected vertices belong to the equivalence classes of an exchangeable equivalence relation. Therefore a decomposition into maximal complete subgraphs may be considered, and, by a straight application of Kingman’s theory of exchangeable random partitions, distributional results for such decomposition may be obtained. We study the case in which vertices colors are a sample from a Dirichlet process. The main motivation being that, under such prior, exchangeable colored random graphs turn out to be an example of logarithmic decomposable combinatorial structure (Arratia et al., 2003).

Many combinatorial structures decompose naturally into components, so that given an instance of size $n$, the most basic description reports only the full component spectrum, i.e. the vector $C_1(n), C_2(n), \ldots, C_n(n)$, where for every $i$, $C_i(n)$ counts the number of components of size $i$. Examples are given by set partitions, whose blocks are the equivalence classes of an arbitrary equivalence relation, cycles decompositions of random permutations, and random graphs decompositions into connected components. When a structure is chosen at random according to some given probability distribution from the set of all structures of size $n$, the vector of components counts can be viewed as a stochastic process which provides both the number and the sizes of the component elements. Component counts of a large class of randomly chosen decomposable combinatorial structures have joint distribution determined by the conditioning relation

$$
\mathcal{L}(C_1(n), \ldots, C_n(n)) = \mathcal{L}(Z_1, \ldots, Z_n | T_n = n)
$$

where $(Z_i, i \geq 1)$ are independent random variables over $Z_+ = \{0, 1, 2, \ldots\}$ and $T_n = \sum_{i=1}^{n} i Z_i$. If, in addition, the $Z_i$ are such to satisfy

$$
\lim_{i \to \infty} i P(Z_i = 1) = \theta = \lim_{i \to \infty} i E Z_i
$$

for some $\theta > 0$, then decomposable combinatorial structures are said to be logarithmic. The reason is that they share the additional property that the expected number of components grows logarithmically with size $n$, $E(K_{0n}) \sim \theta \log n$, where $K_{0n} = \sum_{j=1}^{n} C_j(n)$. (See Arratia et al. (2003) for a comprehensive treatment of the subject).

The simplest example of such a construction is that in which the $Z_i$ are Poisson-distributed with means exactly $\theta/i$. In this case the $C_i(n)$ have as joint distribution the Ewens Sampling Formula of parameter $\theta$ (Ewens, 1972).

It can be shown that classical Erdős random graphs are not logarithmic, since $E(K_{0n})/\log n \to 0$ as $n \to \infty$. Here we show that random graphs with exchangeable
hidden colors, under Dirichlet prior, as an easy consequence of known results about sampling from a Dirichlet process (Antoniak, 1974) provide an example of logarithmic combinatorial random graphs.

The paper is organized as follows. In Section 2 we define the class of ECRG driven by a complete graphs and recall some known results on Kingman’s theory of exchangeable random partitions. In Section 3 we present our main results on cliques decomposition. Finally in Section 4 we discuss some remarks and suggestions for further developments.

2 Preliminaries

Let \( \Gamma = (V, E) \) be a finite graph, of vertex set \( V \), and edges set \( E \). An edge joining the vertices \( i \) and \( j \) is denoted by \( ij \) or \( ji \), without distinction. The number of vertices in \( \Gamma \) is called the order of \( \Gamma \) and it is denoted by \( |\Gamma| \). The number of edges in \( \Gamma \) is called the size of \( \Gamma \) and it is denoted by \( e(\Gamma) \). A graph is a subgraph of \( \Gamma \) if its vertex set and its edge set are subsets of \( V(\Gamma) \) and \( E(\Gamma) \), respectively.

Let \( G \) be the set of graphs with vertex set \( V \) and whose edge set is a subset of \( E \):

\[
G = \{(V, E') : E' \subset E\},
\]

and let \( 2^G \) be the power sigma-algebra of \( G \). Any measurable function \( G \) from a probability space \((\Omega, F, P)\) to the measurable space \((G, 2^G)\) is called a random graph. The probability distribution of \( G \) is the measure induced by \( P \) through \( G \).

To introduce the class of random graphs with hidden colors, let \((S, S)\) be a standard Borel space, named color set. Consider the probability space \((\Omega, F, P)\), where

\[
\Omega = S^\infty = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_i \in S\},
\]

\( F \) is the cylinder sigma-algebra on \( \Omega \) and \( P \) is a probability measure on \( F \). Let \( G : S^\infty \to G \) be defined by

\[
G(\omega) = (V, \{ij \in E : \omega_i = \omega_j\}).
\]

Then \( G \) is a \( F/2^G \)-measurable function.

**Definition 1.** The random element \( G \) defined in (1) is called a random graph with hidden colors.

The coordinate random variables \( X_1, X_2, \ldots \) on \((\Omega, F, P)\), defined by \( X_i(\omega) = \omega_i \), are called the vertices colors. By the previous construction, a random graph with hidden colors can be thought to be obtained from a fixed graph \( \Gamma \), by coloring its vertices according to some law and then deleting the edges with different colors at the endvertices. If the probability distribution of the vertices colors is exchangeable, the random graph is called a random graph with exchangeable hidden colors.
**Definition 2.** Let \( \Omega \) and \( F \) be defined as above and let \( P \) be an exchangeable probability measure on \( F \). A random element \( G \) defined as in (1) is called a random graph with exchangeable hidden colors (ECRG).

If \( G \) is an ECRG, then the vertices colors \( X_1, X_2, \ldots \) form a sequence of exchangeable random variables. In particular, if there exists a probability measure \( P_0 \) on \( S \) such that \( P = P_0^\infty \), then the vertices colors are independent, identically distributed random variables. Random graphs with independent and identically distributed vertices colors can be seen as a particular case of the random randomly colored graphs, introduced by Penman (1998) and by Söderberg (2003), independently.

Now recall that if \([n]\) is a finite set of fixed vertices \( \{1, \ldots, n\} \), the complete graph on \( n \) vertices, \( K_n \), is the graph with vertex set \( V = [n] \), and edge set \( E = \{ij; \forall i, j \in V\} \), of size \( c(K_n) = (\binom{n}{2}) \), hence \( K_n = ([n], [n]^2) \).

**Definition 3.** Let \([n]\) be a finite set of vertices, let \( \Gamma = K_n \), be the complete graph on \( n \) vertices, then for every \( n \), \( G_n : S^\infty \rightarrow \mathcal{G}_n \), defined by

\[
G_n(\omega) = ([n], \{ij \in [n]^{(2)} : \omega_i = \omega_j\})
\]

is called a random graph with exchangeable hidden colors driven by a complete graph.

Now recall that if \( p_n \) is the partition set, the (finite) set of all partitions of \([n]\), \( |p_n| \) the cardinality of the partition set is given by the Bell number \( B_n = \sum_{k=1}^{n} \binom{n}{k} \) where \( \binom{n}{k} \) is the Stirling number of the second kind. Then the \( \sigma \)-field on \( p_n \) is \( 2^{p_n} \), the power set of the partition set, the collection of all subsets of \( p_n \). A random element of \( p_n \) is a random partition.

If \( P \) is exchangeable on \( S^\infty \) there exists a random probability measure \( P_0 \) on \( S \) such that the coordinate r.v.’s \( X_i | P_0 \) are i.i.d. (\( P_0 \)). Let \( \Pi_\infty \) be a random partition of \( \mathbb{N} \) generated by sampling from \( P_0 \), i.e. by the following exchangeable equivalence relation

\[
i \approx j \iff X_i(\omega) = X_j(\omega),
\]

that is to say, two positive integers \( i \) and \( j \) are in the same block of \( \Pi_\infty \) if and only if \( X_i = X_j \). Formally \( \Pi_\infty \) is identified by the sequence \( (\Pi_n) \) where \( \Pi_n \) is the restriction of \( \Pi_\infty \) to the finite set \([n] = \{1, \ldots, n\} \):

\[
\Pi_n(\omega) = \{(A_1, \ldots, A_k) : (i, j) \in A_h \iff X_i(\omega) = X_j(\omega); i, j = 1, \ldots, n; h = 1, \ldots, k\}
\]

and the blocks \( A_h, h = 1, \ldots, k \), are listed in order of appearance, (i.e. in increasing order of their least element). Then \( \Pi_n : \Omega \rightarrow p_n \), is \( F/2^{p_n} \)-measurable. By the following result, due to Kingman, \( \Pi_n \) is an exchangeable random partition of \([n]\).

**Theorem 1.** (Kingman, 1978) The distribution of \( \Pi_n \) is such that for each particular partition \( \{A_1, \ldots, A_k\} \) of \( \{1, 2, \ldots, n\} \), with \( |A_i| = n_i \) for \( 1 \leq i \leq k \), where \( n_i \geq 1 \)
and \( \sum_{i=1}^{k} n_i = n \)

\[
P(\Pi_n = \{A_1, \ldots, A_k\}) = p(n_1, \ldots, n_k)
\]

for some symmetric function \( p \) of sequence of \( k \)-tuples of non-negative integers with sum \( n \), called the exchangeable partition probability function (EPPF) of \( \Pi_\infty \).

Now let \( P_i \) denotes the size of the \( i \)th largest atom of \( P_0 \). If \( P_0 \) is a random discrete distribution, then \( \sum_i P_i = 1 \) almost surely. Assume for simplicity that \( P_i > 0 \) for all \( i \) almost surely, then \( \Pi_\infty \) is said to have proper frequencies \( (P_i) \). In that case, according to Kingman’s representation, the distribution of \( \Pi_n \) is determined by the joint distribution of the sizes of the ranked atoms \( (P_i) \) of \( P_0 \) by the formula:

\[
p(n_1, \ldots, n_k) = \sum_{j_1, \ldots, j_k} E \prod_{i=1}^{k} P_{j_i}^{n_i}
\]

where \( j_1, \ldots, j_k \) ranges over all permutations of \( k \) positive integers.

The previous formula sets up a bijection between probability distributions for an infinite exchangeable random partition, as specified by an infinite EPPF, and probability distributions of \( (P_i) \), the ranked atoms of \( P_0 \), on the set

\[
\nabla := \{(p_1, p_2, \ldots) : p_1 \geq p_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} p_i \leq 1\}
\]

of ranked sub-probability on \( \mathbb{N} \).

We are now in a position to state the following alternative definition of an ECRG driven by a complete graph.

**Definition 4.** Let \( [n] = \{1, \ldots, n\} \), let \( \Pi_n : \Omega \to \mathbb{P}_n \) as defined in (3), be an exchangeable random partition of \( [n] \). Let \( G_n : \mathbb{P}_n \to \mathcal{G}_n \) be defined by

\[
G_n(\{A_1, \ldots, A_k\}) = \{([n], E'), E' \subset [n]^{[2]}, E' = \cup_{j=1}^{k} E_j, E_j = (il : (i, l) \in A_j)\}
\]

Then \( G_n \) is a \( 2^{\mathbb{P}_n}/2^{\mathcal{G}_n} \)- measurable function and it is called an ECRG driven by a complete graph.

In the following we will need another basic definition in graph theory, that here we recall (see e.g. Bollobas, 1998).

**Definition 5.** (Clique) Given a graph \( \Gamma \) on \( n \) vertices, a clique \( KQ \) is a maximal totally connected subgraph of \( \Gamma \), i.e. every vertices in \( KQ \) is directly connected with every other vertex in \( KQ \) and \( KQ \) contains all such vertices (i.e. it is maximal).

Let \( \mathcal{M} \) be the set of all subsets of \( m \) vertices of \( [n] \), hence \( |\mathcal{M}| = \binom{n}{m} \) and let \( \alpha \in \mathcal{M} \). Consider the random subgraph of \( G_n \) spanned by \( \alpha \), i.e. the restriction of \( G_n \) to the vertices \( [\alpha] \):

\[
G_{n,\alpha}(\omega) = ([\alpha], \{ij \in [\alpha]^{[2]} : \omega_i = \omega_j\})
\]
Let $\bar{\alpha} = [n] \setminus \alpha$, hence $|\bar{\alpha}| = (n - m)$. Call $|\mathcal{M} + 1|$ the set of all subsets of $m + 1$ vertices of $[n]$, $|\mathcal{M} + 1| = \binom{n}{m+1}$, and let $\mathcal{M}' \subset \mathcal{M} + 1$, be the set of all subsets of $m + 1$ vertices, such that $m$ vertices are in $\alpha$,

$$\mathcal{M}' = \{(v_1, \ldots, v_{m+1}) : \{v_1, \ldots, v_m\} = \alpha, v_{m+1} \in \bar{\alpha}\}$$

so that $|\mathcal{M}'| = (n - m)$.

It is easy to see that $\{G_n,\alpha\} = \{G_n,\alpha = KQ_{\alpha}\} \cap \bigcap_{\alpha' \in \mathcal{S}' \setminus \mathcal{S}} \{G_n,\alpha' \neq K_{\alpha'}\}$, i.e. $G_n,\alpha$ is a clique if $G_n,\alpha$ is complete and $G_n,\alpha'$ is not complete for every $\alpha' \in \mathcal{M}'$.

### 3 Main results

First we show that an ECRG driven by a complete graph naturally decomposes into cliques, whose vertices are the elements of the blocks of an exchangeable random partition.

By the previous section, if $G_n$ is defined as in (7), the random subgraph restricted to $[\alpha] = \{i_1, \ldots, i_m\}$, is defined by

$$G_{n,\alpha}(\pi^k_n) = \{([\alpha], E'), E' \subset [\alpha]^{(2)}, E' = \cup_{j=1}^k E_j, E_j = \{il : (i, l) \in A_j, i, l = i_1, \ldots, i_m\}\}$$

for $\pi^k_n = \{A_1, \ldots, A_k\}$. Now $G_{n,\alpha}$ is complete if and only if $E' = [\alpha]^{(2)} = \{(il) : i, l = i_1, \ldots, i_m\}$, hence if and only if $[\alpha]^{(2)} = \cup_{j=1}^k E_j = \cup_{j=1}^k \{il : (i, l) \in A_j, i, l = i_1, \ldots, i_m\}$. Hence if and only if $\exists j : [\alpha] \subseteq A_j$. Moreover $G_{n,\alpha}$ is a clique if $G_{n,\alpha'} \neq K_{\alpha'}$, $\forall \alpha' \in \mathcal{M}'$, i.e.

$$\{G_{n,\alpha}(\pi^k_n) = KQ_{\alpha}\} = \left(\bigcup_{j=1}^k \{[\alpha] \subseteq A_j\}\right) \cap \left(\bigcap_{\alpha' \in \mathcal{M}' \setminus \mathcal{S}} \left[\bigcap_{j=1}^k \{[\alpha'] \supset A_j\}\right]\right)$$

Therefore $G_{n,\alpha}$ is a clique if and only if $\exists j : [\alpha] = A_j$. Since $\{A_1, \ldots, A_k\}$ is a partition of $[n]$, a unique $j$ such that $[\alpha] = A_j$ may exists, so that the following proposition, that we state without proof, establishes the obvious decomposition.

**Proposition 1.** (Cliques decomposition) Let $G_n$ be an ECRG on $[n]$ driven by a complete graph, whose vertices colors are a sample from a random discrete distribution $P_0$. Then there exists a random partition $\{\alpha_1, \ldots, \alpha_k\}$ of $[n]$, such that the following decomposition holds:

$$G_n(\omega) = \bigcup_{j=1}^k G_{n,\alpha_j}(\omega) \quad (8)$$
where \([n] = \bigcup_{j=1}^{k} [\alpha_j]\), and, for every \(j\), \(G_{n,\alpha_j}(\omega) = \{([\alpha_j], E_j), E_j = (i : i, l \in [\alpha_j])\}\). Moreover, for each \(j\), \(G_{n,\alpha_j}\) is a clique, and \([\alpha_j] = A_j\), where \(A_j\) are the classes of equivalence of the random exchangeable partition \(\Pi_n\) generated by sampling from \(P_0\).

**Remark 1.** Many different kinds of random graphs decompose in connected components, (see Example 2.3 in Arratia et al., 2003). By Proposition 1, an ECRG driven by a complete graph turns out to be a random decomposable combinatorial structure, whose components are cliques, maximal complete subgraphs. By the construction illustrated so far, cliques result by the operation of random coloring the vertices according to an exchangeable probability distribution. Hence the vertices set of each clique coincide with the elements set of a block of an exchangeable random partition.

It is known that given an instance of size \(n\), the most basic description of a decomposable combinatorial structure, only reports the component spectrum, i.e. the vector \(C^{(n)} = (C_1^{(n)}, C_2^{(n)}, \ldots, C_n^{(n)})\), specifying how many components there are of size one, two, three and so on, with the obvious constrain that \(\sum_{i=1}^{n} jC_j^{(n)} = n\).

For a random combinatorial structure, when a structure is chosen according to some given probability distribution from the set of all structures of size \(n\), the description is given in terms of the joint law of the component spectrum.

In our framework, the component spectrum is given by the random vector counting how many cliques there are of order one, two, three and so on, in the decomposition (8). Now, for every \(m = 1, \ldots, n\), let

\[
\xi_m(G_{n,\alpha_j}) = \begin{cases} 
1 & \text{if } |\alpha_j| = m \\
0 & \text{if } |\alpha_j| \neq m
\end{cases}
\]

be the indicator function of a clique of order \(m\), then the number of cliques of order \(m\) in \(G_n\), is given by the random variable:

\[
C_m(G_n) = \sum_{\alpha_j \in \{\alpha_1, \ldots, \alpha_k\}} \xi_m(G_{n,\alpha_j}),
\]

so that the **clique counts vector** is given by the random vector

\[
C(G_n) = (C_1(G_n), C_2(G_n), \ldots, C_n(G_n)),
\]

taking values \(c_1, \ldots, c_n\), where for each \(m\), \(c_m \in \mathbb{Z}_+\) and \(\sum_{m=1}^{n} mc_m = n\).

Now for a measure \(\mu\) on \(S\) with \(0 < \mu(S) < \infty\) let \(\theta = \mu(S)\) and \(\nu = \mu/\theta\). So \(\theta > 0\), \(\nu\) is a probability distribution on \(S\), and \(\mu = \theta\nu\). Let the vertices colors, \(X_1, X_2, \ldots\) be a sample from a Dirichlet process with parameter \(\theta\nu\), for \(\theta > 0\) and \(\nu\) a diffuse probability measure, \(X_i|P_0\) i.i.d. \(\sim P_0\), with \(P_0 \sim \mathcal{D}(\mu)\).

**Theorem 2.** (Dirichlet colored random graphs.) Let \(G_n\) be a ECRG on \([n]\) driven by a complete graph. Let the vertices colors be a sample from \(P_0\), with \(P_0 \sim \mathcal{D}(\theta\nu)\),
then the joint distribution of the clique counts vector $C_1(G_n), C_2(G_n), \ldots, C_n(G_n)$, is given by the Ewens Sampling Formula of parameter $(\theta)$,

$$P(C_m(G_n) = c_m, 1 \leq m \leq n) = \frac{n!}{\theta^{(n)}} \prod_{m=1}^{n} \left( \frac{\theta}{m} \right)^{c_m} \frac{1}{c_m!},$$

for $c_1, \ldots, c_n$ such that $c_m \in \mathbb{Z}_+$, for every $m = 1, \ldots, n$, and $\sum_{m=1}^{n} mc_m = n$.

**Proof:** By construction, the clique counts vector $C(G_n)$ turns out to be the vector counting the sizes of the blocks in an exchangeable random partition generated by sampling from the random probability measure $P_0$. The thesis follows by a straight application of a known result (Antoniak, 1974), based on the Blackwell-MacQueen description of sampling from a Dirichlet prior, which states that the distribution of the partition induced by a sample from a Dirichlet process $(\theta \nu)$, is given by the Ewens Sampling Formula of parameter $\theta$.

**Remark 2.** The Ewens Sampling Formula (Ewens, 1972) originally arose in population genetics, where the parameter $\theta$ is a mutation rate. It has been established for many different models; see Arratia et al. (2003), Johnson, Kotz and Balakrishnan (1997, Ch. 41). For each $n = 1, 2, \ldots$ and $\theta > 0$, it gives the joint distribution of the component counts for a random permutation of $n$ objects, chosen with probability biased by $\theta^{K_0_n}$, where $K_0_n$ is the number of cycles. It provides the simplest example of decomposable combinatorial structure whose component counts $C_1^{(n)}, \ldots, C_n^{(n)}$, have joint distribution determined by the **Conditioning Relation**:

$$\mathcal{L}(C_1^{(n)}, \ldots, C_n^{(n)}) = \mathcal{L}(Z_1, \ldots, Z_n | T_n = n)$$

where $(Z_i, i \geq 1)$ are independent random variables over $\mathbb{Z}_+$ and $T_n = \sum_{i=1}^{n} iZ_i$, for $Z_i$ Poisson random variables with means $EZ_i = \theta / i$, $i = 1, 2, \ldots, n$.

**Remark 3.** The asymptotic theory of decomposable combinatorial structures is the subject of a recent book by Arratia et al. (2003). They consider randomly chosen combinatorial structures of total size $n$, whose component counts $C_1^{(n)}, \ldots, C_n^{(n)}$, satisfy, in addition to the **Conditioning Relation**, which is common to many classical examples, the **Logarithmic Condition**:

$$\lim_{i \to \infty} iP(Z_i = 1) = \theta = \lim_{i \to \infty} iEZ_i$$

for some $\theta > 0$. Logarithmic structures share the additional property that the expected number of components grows logarithmically with the size $n$ as $n \to \infty$. It is easy to check that the Ewens Sampling Formula satisfies the logarithmic condition. Thus ECRG’s driven by a complete graph, under a Dirichlet process, are logarithmic combinatorial structures, as we state in the following theorem.
Theorem 3. (Logarithmic combinatorial random graphs.) Let \((G_n)_n\) be a sequence of ECRG’s driven by the sequence \((K_n)_n\) of complete graphs on \(n\) vertices. For each \(n\), let \(\{G_{n,\alpha_1}, \ldots, G_{n,\alpha_k}\}\) be the cliques decomposition, and let

\[
KQ_{0n}(G_n) = \sum_{m=1}^{n} \sum_{\alpha_j \in \{\alpha_1, \ldots, \alpha_k\}} \xi_m(G_{n,\alpha_j})
\]

be the total number of cliques in the decomposition. If the vertices colors are a sample from \(P_0\), with \(P_0 \sim D(\theta \nu)\), then

\[
E(KQ_{0n}) \sim \theta \log(n)
\]

i.e. the expected number of cliques grows logarithmically with the size \(n\).

Proof: By construction, the number of cliques in \(G_n\) turns out to be the number of blocks in an exchangeable random partition of \([n]\) generated by sampling from \(D(\theta \nu)\). The thesis follows by a result in Korwar and Hollander (1973), which states that, if \(D_n\) is the number of distinct observations in a sample \(X_1, \ldots, X_n\) from a Dirichlet process with parameter \(\mu\) a non atomic measure, then \(D_n / \log n \overset{a.s.}{\rightarrow} \mu(S)\) as \(n \to \infty\).

Now recall that given a graph \(H\), the number of vertices is called the order of the graph and it is denoted by \(|H|\). Consider a ECRG driven by a complete graph on \(n\) vertices. Let \(G_{n,\alpha_1}, G_{n,\alpha_2}, \ldots\) be the cliques in the decomposition (8) arranged in the decreasing order of their orders. The following theorem states that the vector of the ranked normalized cliques orders in a Dirichlet colored random graph, has a limiting Poisson-Dirichlet (\(\theta\)) distribution, (Kingman, 1975).

Theorem 4. (Ranked normalized cliques orders.) Let \((G_n)_n\) be a sequence of ECRG’s driven by a sequence \((K_n)_n\) of complete graphs on \(n\) vertices. Let the vertices colors be a sample from \(P_0\), where \(P_0 \sim D(\theta \nu)\), then

\[
\frac{|G_{n,\alpha_1}|}{n}, \frac{|G_{n,\alpha_2}|}{n}, \ldots \overset{d}{\rightarrow} PD(\theta)
\]

as \(n \to \infty\), where \(PD(\theta)\) is the Poisson-Dirichlet distribution with parameter \(\theta\).

Proof: By the well-known Ferguson’s constructive definition of a Dirichlet process (Ferguson, 1973), if \(P_i\) denotes the magnitude of the \(i\)-th largest atom of \(P_0\), then, almost surely,

\[
P_0 = \sum_{i=1}^{\infty} P_i \delta(\tilde{X}_i)
\]

where the \(\tilde{X}_i\) are i.i.d. \((\nu)\) and the sequence \((P_i)\), with \(P_1 > P_2 > \cdots > 0\) and \(\sum_i P_i = 1\) almost surely, has the Poisson-Dirichlet distribution with parameter \(\theta\). By construction, the decreasing normalized cliques orders coincide with the ranked
normalized sizes of the blocks of an exchangeable random partition generated by sampling from a Dirichlet process with parameter $\theta\nu$. The convergence to the Poisson-Dirichlet distribution was proved by Kingman (1977). (See also Aldous, 1985, Th. 11.14, which states the same result about the decreasing rearrangement of the normalized sizes of the blocks in the Ewens sampling partition structure).

It is always possible to assume an arbitrary ordering for the graph’s vertices, so that, given the sequence $X_1, X_2, \ldots$ of the vertices colors, the order of appearance of different colors may be considered. Let $X_i$ be the observed color of the $i$-th vertex and let $\hat{X}_j$ be the $j$-th distinct color to appear. Let $G_n,\hat{\alpha}_1, G_n,\hat{\alpha}_2, \ldots$ denotes the random vector of cliques, ordered by their least vertex element. As an easy consequence of the previous theorem, the asymptotic distribution of the normalized cliques orders in order of appearance, is a GEM distribution with parameter $\theta$.

**Theorem 5.** (Normalized cliques orders in order of appearance.) Let $(G_n)_n$ be a sequence of ECRG’s driven by a sequence $(K_n)_n$ of complete graphs on $n$ vertices. Let the vertices colors be a sample from $P_0$, where $P_0 \sim D(\theta\nu)$, then

$$\frac{|G_n,\hat{\alpha}_1|}{n}, \frac{|G_n,\hat{\alpha}_2|}{n}, \ldots \xrightarrow{d} \text{GEM}(\theta)$$

as $n \to \infty$.

**Proof:** It is easy checked that the vector $\frac{|G_n,\hat{\alpha}_1|}{n}, \frac{|G_n,\hat{\alpha}_2|}{n}, \ldots$ is a size-biased permutation of $\frac{|G_n,\alpha_1|}{n}, \frac{|G_n,\alpha_2|}{n}, \ldots$. Since it is known that the size-biased permutation of the Poisson-Dirichlet distribution with parameter $\theta$ is GEM $(\theta)$, the thesis follows.

## 4 Final remarks

Random graphs with exchangeable hidden colors have been introduced very recently. This paper arises from observing that vertices remaining directly connected after the act of coloring, are exactly the classes of equivalence of an exchangeable equivalence relation. The results we obtained are an easy application of well-known results about the theory of exchangeable random partitions. We restricted to the simplest case, the one arising when the ECRG driven by a complete graph is governed by a Dirichlet process. Nevertheless, it would be interesting to investigate the distributional behaviour of the cliques decomposition when the vertices colors are a sample from more general random discrete distributions. An appealing and obvious generalization would be to the case in which the vertices colors are a sample from a species sampling model (Pitman, 1996, 2002). In fact it is known that from such models arises a more general theory of exchangeable random partition extending Blackwell-MacQueen construction of a Dirichlet process.

It would be also interesting to study the possibility that ECRG’s driven by a complete graph, may represent a limit model for more general families of random
colored graphs.

**Acknowledgements.** I wish to thank Sandra Fortini for her constant support, encouragement and helpful suggestions.

**References**


