LA-UR- 10-01561

Approved for public release; distribution is unlimited.

Title: COMPONENT EVOLUTION IN GENERAL RANDOM INTERSECTION GRAPHS

Author(s): Milan Bradonjic
Aric Hagberg
Nick Hengartner
Allon G. Percus

Intended for: AOFA 10
Vienna Austria - 6/25-7/02/2010

Los Alamos
NATIONAL LABORATORY
EST. 1943

Component Evolution in General Random Intersection Graphs

Milan Bradonjić [*], Aric Hagberg [†], Nick Hengartner [‡], Allon G. Percus [§]

Abstract

We analyze component evolution in general random intersection graphs (RIGs) and give conditions on existence and uniqueness of the giant component. Our techniques generalize the existing methods for analysis on component evolution in RIGs. That is, we analyze survival and extinction properties of a dependent, inhomogeneous Galton-Watson branching process on general RIGs. Our analysis relies on bounding the branching processes and inherits the fundamental concepts from the study on component evolution in Erdős-Rényi graphs. The main challenge becomes from the underlying structure of RIGs, when the number of offsprings follows a binomial distribution with a different number of nodes and different rate at each step during the evolution. RIGs can be interpreted as a model for large randomly formed non-metric data sets. Besides the mathematical analysis on component evolution, which we provide in this work, we perceive RIGs as an important random structure which has already found applications in social networks, epidemic networks, blog readership, or wireless sensor networks.

Keywords: Random graphs, branching processes, probabilistic methods, random generation of combinatorial structures, stochastic processes in relation with random discrete structures.

1 Introduction

Bipartite graphs, consisting of two sets of nodes with edges only between nodes in opposite sets, are often the natural representation for classification of objects where each objects has a set of properties [10]. Collaboration graphs also are common examples of bipartite graphs where, for example, the sets of scientists and research papers or actors and movies form the two sets [23, 15]. In general most social networks can be cast as bipartite graphs since they are built from sets of individuals connected to sets of attributes such as membership of a club or organization, work colleagues, or fans of the same sports team. For example, simulations of epidemic spread in human populations are often performed on networks constructed from bipartite graphs of people and the locations they visit during a typical day [11]. Social networks, however, are not the only networks with bipartite structure, any set of relations between objects and properties form a bipartite graph. For example, the relation between network nodes and keys in an implementation of a secure wireless network forms a bipartite network [5].

Modeling social networks or object classification networks remains a challenge. The well-studied Erdős-Rényi model, $G_{n,p}$, successfully used for average case analysis of algorithm performance, does not well represent many randomly formed networks, such as social or collaboration networks. For example, $G_{n,p}$ does not capture the typical scale-free degree distribution of most real-world networks [3]. More realistic degree distributions can be achieved by the configuration model [17] or expected degree model [7], but those still fail to capture other common properties of social networks such as high number of triangles (or cliques) and strong degree-degree correlation [16, 1]. Extension of the configuration model to specify the degree of both bipartite sets remedied some of these problems [13].

[*]Theoretical Division, and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, milan@lanl.gov,

[†]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, hagberg@lanl.gov,

[‡]Statistical Sciences Group Los Alamos National Laboratory, NM 87545, USA, nickh@lanl.gov,

[§]School of Mathematical Sciences Claremont Graduate University, Claremont, CA 91711, USA, allon.percus@cgu.edu.

Another closely related alternative set of models, are random intersection graphs which were first introduced in [22, 14]. Any undirected graph can be represented as an intersection graph [9]. The simplest version is the "uniform" RIG, $G(n, m, p)$ where each of the n nodes in the graph are connected with the same probability p to a random subset of m elements in the attribute set. Then, two nodes in the graph are connected, if and only if they are connected to at least one element in the attribute set. A more general model, and the one which we study in this work, is the general RIG, $G(n, m, \boldsymbol{p})$, where the set of probabilities $\boldsymbol{p}$, which connects nodes to attributes, is not uniform, but rather has unique values for each attribute [19, 18]. This general model has only recently been developed and only a few results, such as expander properties, cover time, and the existence and efficient construction of large independent sets, have been studied [19, 18, 20].

In this paper we analyze the component evolution in general random intersection graphs which generalizes the results for uniform random intersection graphs [4]. Two other special cases of the general random intersection graph model, with a specific overlap threshold controlling the connectivity of the nodes, were analyzed in [5]. Our analysis also generalizes the component evolution methods used on $G_{n,p}$ [2, 21]. Specifically, in order to study the component evolution we analyze the branching processes (the behavior of a breath-first search algorithm) when the number of offsprings follows a binomial distribution with a different number of nodes and different rates at each step during the evolution. This approach is a generalization of the case when the number of offsprings of the nodes at the same level of the branching tree follow the same distribution [6]. The main challenge becomes from the underlying structure of RIGs, when the number of offsprings follows a binomial distribution with a different number of nodes and different rate at each step during the evolution.

2 Model and previous work

In this paper we will use the notation for random intersection graphs introduced in [5].

Model. There are two sets: the set of nodes $V = \{1, 2, \dots, n\}$ and the set of attributes $W = \{1, 2, \dots, m\}$. Every node $v \in V$ is assigned a random set of attributes $S(v) \subseteq W$. Then two nodes $u, v \in V$ are connected if, and only if, $|S(u) \cap S(v)| \geq s$, for a given integer $s \geq 1$. We consider the general intersection graph $G(n, m, \boldsymbol{p})$, introduced in [19, 18], with $\boldsymbol{p} = \{p_w\}_{w \in W}$, where $p_w \in (0, 1)$ [1], and, for all $v \in V$ and $w \in W$, $\mathbb{P}[w \in S(v)] = p_w$. We consider the general case when p_w are not necessarily the same and for simplicity we fix $s = 1$.

The component evolution of the uniform model $G(n, m, p)$ was analyzed by Behrisch in [4], for the case when the scaling of nodes and attributes is $m = n^\alpha$, with $\alpha \neq 1$ and $p^2 m = c/n$. Theorem 1 in [4] states that the size of the largest component $\mathcal{N}(G(n, m, p))$ in RIG satisfies (i) $\mathcal{N}(G(n, m, p)) \leq \frac{9}{(1-c^2)} \log n$, for $\alpha > 1, c < 1$, (ii) $\mathcal{N}(G(n, m, p)) = (1 + o(1))(1 - \rho)n$, for $\alpha > 1, c > 1$, (iii) $\mathcal{N}(G(n, m, p)) \leq \frac{10\sqrt{c}}{(1-c^2)} \sqrt{\frac{n}{m}} \log m$, for $\alpha < 1, c < 1$, (iv) $\mathcal{N}(G(n, m, p)) = (1 + o(1))(1 - \rho)\sqrt{cmn}$, for $\alpha < 1, c > 1$, where ρ is the solution in $(0, 1)$ of the equation $\rho = \exp(c(\rho - 1))$.

The component evolution for the case $s \geq 1$ in the relation $|S(u) \cap S(v)| \geq s$ is considered in [5], where the following two RIG models are analyzed: (1) $G_s(n, m, d)$ model, where $\mathbb{P}[S(v) = A] = \binom{m}{d}^{-1}$ for all $A \subseteq W$ on d elements, for a given d; (2) $G'_s(n, m, p)$ model, where $\mathbb{P}[S(v) = A] = p^{|A|}(1-p)^{m-|A|}$ for all $A \subseteq W$. In light of results of [4], it has been shown in [5], that for $d = d(n), p = p(n), m = m(n), n = o(m)$, where s is a fixed integer, and $d^{2s} \sim cm^s s!/n$, the largest component in $G_s(n, m, d)$ satisfies: (i) $\mathcal{N}(G_s(n, m, d)) \leq \frac{9}{(1-c^2)} \log n$, for $c < 1$, (ii) $\mathcal{N}(G_s(n, m, d)) = (1 + o(1))(1 - \rho)n$, for $c > 1$, in the case when $n \log^n = o(m)$ for $s = 1$ and $n = o(m^{s/(2s-1)})$ for $s \leq 2$. The same results for the giant component in $G_s(n, m, p)$ still hold for the case when $p^{2s} = cs!/m^s n$ and $n = o(m^{s/(2s-1)})$ [5].

Both $G_s(n, m, d)$ and $G'_s(n, m, p)$ are special cases of a more general class studied in [12], where the number

[1] We can eliminate the cases $p_w = 0$ and $p_w = 1$, since either none or all of the nodes would have the attribute w and we are interested in graphs without isolated nodes or completely connected nodes. Note that the p_w's do not sum up to 1.

of attributes of each node is assigned randomly as in the bipartite configuration model. That is, for a given probability distribution $(P_0, P_1, \dots, P_m)$, we have $\mathbb{P}[|S(v)| = k] = P_k$ for all $0 \leq k \leq m$, and moreover given the size k, all of the sets $S(v)$ are equally probable, that is for any $A \subseteq W$, $\mathbb{P}[S(v) = A : |S(v)| = k] = \binom{m}{k}^{-1}$. That is, we see that $G_s(n, m, d)$ is equivalent to the model of [12] with the delta-distribution, where the probability of the d-th coordinate is 1, while $G'_s(n, m, d)$ is equivalent to the model of [12] with the $\mathrm{Bin}(m, p)$ distribution. Comparing $G(n, m, \boldsymbol{p})$, the general RIG model, with the model of [12], it follows that general RIG model does not perform a "uniform sampling", as the model of [12], which we explain in (6), Section 3. To complete the picture of previous work, it was shown that when $n = m$ a of probabilities $\boldsymbol{p} = \{p_w\}_{w \in W}$ can be chosen to tune the degree and clustering coefficient of the graph [8].

3 Mathematical preliminaries

The edges in RIG are not independent. To see this, consider three distinct nodes $u, v, w \in V$. Conditionally on the set $S(w)$, the random sets $S(u)$ and $S(v)$ are conditionally independent, since given $S(w)$, the sets $S(u) \cap S(w)$ and $S(v) \cap S(w)$ are mutually independent. The latter implies conditional independence of the events $\{u \sim w \mid S(w)\}, \{v \sim w \mid S(w)\}$, and hence

$$\mathbb{P}[u \sim w, v \sim w \mid S(w)] = \mathbb{P}[u \sim w \mid S(w)]\mathbb{P}[v \sim w \mid S(w)]. \tag{1}$$

However, the latter does not imply imply independence of the events $\{u \sim w\}$ and $\{v \sim w\}$ since

$$\begin{aligned} \mathbb{P}[u \sim w, v \sim w] &= \mathbb{E}[\mathbb{P}[u \sim w, v \sim w \mid S(w)] \\ &= \mathbb{E}\left[\mathbb{P}[u \sim w \mid S(w)]\mathbb{P}[v \sim w \mid S(w)]\right] \\ &\neq \mathbb{P}[u \sim w]\mathbb{P}[v \sim w]. \end{aligned} \tag{2}$$

Furthermore, the conditional pairwise independence (1) does not extend to three or more nodes. Indeed, conditionally on the set $S(w)$, the sets $S(u) \cap S(v)$, $S(u) \cap S(w)$, and $S(v) \cap S(w)$ are not mutually independent, and hence neither are the events $\{u \sim v\}$, $\{u \sim w\}$, and $\{v \sim w\}$, that is,

$$\mathbb{P}[u \sim v, u \sim w, v \sim w \mid S(w)] \neq \mathbb{P}[u \sim v \mid S(w)]\mathbb{P}[u \sim w \mid S(w)]\mathbb{P}[v \sim w \mid S(w)]. \tag{3}$$

We now provide two identities that we will use throughout this paper. For any $w \in W$, let $q_w := 1 - p_w$, and define $\prod_{\alpha \in \emptyset} q_\alpha = 1$.

Claim 1 *For any node $u \in V$ and fixed set $A \subseteq W$,*

$$\mathbb{P}[S(u) \cap A = \emptyset] = \prod_{\alpha \in A} (1 - p_\alpha) = \prod_{\alpha \in A} q_\alpha. \tag{4}$$

Proof Write

$$\mathbb{P}[S(u) \cap A = \emptyset] = \mathbb{P}[\forall \alpha \in A, \alpha \notin S(u)] = \prod_{\alpha \in A} \mathbb{P}[\alpha \notin S(u)] = \prod_{\alpha \in A} (1 - p_\alpha) = \prod_{\alpha \in A} q_\alpha,$$

which is the desired expression. ■

Claim 2 *For any node $u \in V$, and fixed sets $A, B \subseteq W$,*

$$\mathbb{P}[S(u) \cap A = \emptyset, S(u) \cap B \neq \emptyset] = \Big(\prod_{\alpha \in A} q_\alpha\Big)\Big(1 - \prod_{\alpha \in B \setminus A} q_\beta\Big) = \prod_{\alpha \in A} q_\alpha - \prod_{\beta \in B} q_\beta.$$

Proof The sets A and $B \setminus A$ are disjoint. The result follows from (4). ■

It follows from (4) that for any node $u, v \in V$, $\mathbb{P}[u \sim v | S(v)] = 1 - \prod_{\alpha \in S(v)} q_\alpha$. Taking the expectation over $S(u)$ yields

$$\mathbb{P}[u \sim v] = \sum_{S(u) \subseteq W} \mathbb{P}[S(u)] \Big(1 - \prod_{\alpha \in S(v)} q_\alpha\Big) = 1 - \prod_{w \in W} (1 - p_w^2). \tag{5}$$

To explain non-uniform sampling of the general RIG model, it in general follows that for a given set of probabilities $\{p_w\}_{w \in W}$:

$$\mathbb{P}[S(v) = A \mid |S(v)| = k] = \frac{\mathbb{P}[S(v) = A, |S(v)| = k]}{\mathbb{P}[|S(v)| = k]} = \frac{\prod_{\alpha \in A} p_\alpha \prod_{\beta \notin A} (1 - p_\beta)}{\sum_{A \subseteq W, |A| = k} \prod_{\alpha \in A} p_\alpha \prod_{\beta \notin A} (1 - p_\beta)} \neq \frac{1}{\binom{m}{k}}. \tag{6}$$

4 Branching process on random intersection graphs

Our analysis for the emergence of a giant component is inspired by the approach described in [2]. The idea is to define an auxiliary process related to a breath-first search (BFS) algorithm starting at an arbitrary node $v_0 \in V$, whose stopping time is the size of the component containing v_0. To define this process, we label at each time t the nodes in V as *alive*, *neutral*, or *dead*. Specifically, we initialize the process at time $t = 0$ by labeling all the nodes in V as neutral. Then we pick one node v_0 at random among the neutral node, and label all the nodes connected to v_0 as alive. Finally, label v_0 as dead. At each subsequent times $t \geq 1$, pick one node v_t uniformly at random from the alive nodes and label all the neutral nodes connected to v_t as alive, and v_t as dead.

We describe this process in terms of the random variables (N_t, Y_t, Z_t): N_t the number of remaining neutral nodes at the end of iteration t, Y_t the number of alive nodes at the end of iteration t, and Z_t the number of neutral nodes that become alive in the course of iteration t. These random variables satisfy $Y_0 = 1, N_0 = n - 1$ and the recursion relation

$$Y_t = Y_{t-1} + Z_t - 1, \text{ for } t \geq 1, \tag{7}$$

$$N_t = n - 1 - \sum_{\tau=1}^{t} Z_\tau, \text{ for } t \geq 1. \tag{8}$$

Moreover

$$Y_t - Y_0 = \sum_{\tau=1}^{t} Z_\tau - t, \text{ for } t \geq 1, \tag{9}$$

$$N_t = n - t - Y_t, \text{ for } t \geq 0. \tag{10}$$

Define the stopping time

$$T(v_0) = \inf\{t > 0 : Y_t = 0\}. \tag{11}$$

By construction of the process $\{Y_t\}$, the size of the connected component $C(v_0)$ containing v_0 is $T(v_0)$. Note that the previously defined process on a graph stops at the first t for which $Y_t = 0$. However, we can formally consider the set of the previous equations, for any t, that is even when $Y_t < 0$. This set of equations, we may later call "BFS" equations.

4.1 The probability of exposing neutral nodes during the branching process

As in [5], we denote the cumulative feature set associated to the sequence of nodes $v_0, \ldots, v_t$ from the BFS algorithm by

$$S_{[t]} := \cup_{\tau=0}^{t-1} S(v_\tau). \tag{12}$$

Furthermore we define the history of the BFS at time t to be

$$\mathcal{H}_t = \{v_0, v_1, \ldots, v_t, S_{[0]}, S_{[1]}, \ldots, S_{[t]}\}. \tag{13}$$

In light of Claim 2, the conditional probability r_t of exposing a given neutral node u given the history $\mathcal{H}_t$ is

$$\begin{aligned} r_t &:= \mathbb{P}[u \sim v_t, u \not\sim v_{t-1}, u \not\sim v_{t-2}, \ldots, u \not\sim v_0 | \mathcal{H}_t] \\ &= \mathbb{P}[S(u) \cap S(v_t) \neq \emptyset, S(u) \cap S_{[t-1]} = \emptyset | \mathcal{H}_t] \\ &= \mathbb{P}[S(u) \cap S(v_t) \neq \emptyset, S(u) \cap S_{[t-1]} = \emptyset | S(v_t), S_{[t-1]}] \\ &= \prod_{\alpha \in S_{[t-1]}} q_\alpha - \prod_{\beta \in S_{[t]}} q_\beta \\ &= \phi_{t-1} - \phi_t, \end{aligned} \tag{14}$$

where we set $\phi_t := \prod_{\alpha \in S_{[t]}} q_\alpha$, and define $S_{[-1]} \equiv \emptyset$, $\phi_{-1} \equiv 1$. Observe that this probability is the same for all neutral nodes. Hence the number of neutral nodes becoming alive at time t is, conditionally on the history $\mathcal{H}_t$, a Binomial distributed random variable with parameters N_t and r_t. Formally,

$$Z_t | \mathcal{H}_t \sim \text{Bin}(N_t, r_t). \tag{15}$$

This allows us to describe the distributions of N_t and Y_t in the next lemma.

Lemma 3 *For times $t \geq 1$, the number of neutral and alive nodes satisfies*

$$N_t | \mathcal{H}_{t-1} \sim \text{Bin}\Big(n-1, \prod_{\tau=0}^{t-1}(1-r_\tau)\Big). \tag{16}$$

and

$$Y_t | \mathcal{H}_{t-1} \sim \text{Bin}\Big(n-1, 1-\prod_{\tau=0}^{t-1}(1-r_\tau)\Big) - t + 1. \tag{17}$$

The proof of this lemma requires us to establish the following result first.

Lemma 4 *Let random variables Λ_1, Λ_2 satisfy: $\Lambda_1 \sim \text{Bin}(m, \nu_1)$. Λ_2 given $\Lambda_1 \sim \text{Bin}(\Lambda_1, \nu_2)$. Then marginally $\Lambda_2 \sim \text{Bin}(m, \nu_1\nu_2)$ and $\Lambda_1 - \Lambda_2 \sim \text{Bin}(m, \nu_1(1-\nu_2))$.*

Proof Let $U_1, \ldots, U_m$ and $V_1, \ldots, V_m$ be i.i.d. Uniform$(0,1)$ random variables. Writing

$$\Lambda_1 \stackrel{d}{=} \sum_{j=1}^{m} \mathbb{I}(U_j \leq \nu_1) \quad \text{and} \quad \Lambda_2 | \Lambda_1 \stackrel{d}{=} \sum_{k: U_k < \nu_1} \mathbb{I}(V_k \leq \nu_2),$$

we have that

$$\Lambda_2 \stackrel{d}{=} \sum_{k=1}^{m} \mathbb{I}(U_k \leq \nu_1)\mathbb{I}(V_k \leq \nu_2) \stackrel{d}{=} \sum_{k=1}^{m} \mathbb{I}(U_k \leq \nu_1\nu_2),$$

from which the conclusion follows. ■

Proof (Proof of Lemma 3) We prove the Lemma by induction in t. For $t = 0$, we have $N_0 = n - 1$, and $Z_1 \sim \text{Bin}(N_0, r_0)$, that is, $Z_1 \sim \text{Bin}(n-1, r_0)$. Using (8), it follows that $N_1 = n - 1 - Z_1 \sim \text{Bin}(n-1, r_0)$. Thus the Lemma is true for $t = 0$. Assume that the Lemma is true for some $t \geq 1$,

$$N_t | \mathcal{H}_{t-1} \sim \text{Bin}\big(n-1, \prod_{\tau=0}^{t-1}(1-r_\tau)\big). \tag{18}$$

From (15), we know that $Z_{t+1}|\mathcal{H}_t \sim \text{Bin}(N_t, r_t)$ and (10) implies $N_{t+1} = N_t - Z_{t+1}$. Now from Lemma 4, it follows

$$N_{t+1}|\mathcal{H}_t \sim \text{Bin}\big(n-1, \prod_{\tau=0}^{t}(1-r_\tau)\big). \tag{19}$$

Hence, by mathematical induction, the Lemma holds for any $t \geq 0$. ■

4.2 Expectation of ϕ_t

We now focus on describing the distribution of $\phi_t = \prod_{\alpha \in S_{[t]}} q_\alpha$. Given any sequence $v_0, v_1, v_2, \ldots, v_{n-1}$ enumerating the nodes in V and $w \in W$, let Γ_w denote the first time that a node in that sequence connects to w. We set $\Gamma_w = \infty$ if none of the vertices in V connect to w. That is $\Gamma_w = k$ if $w \notin S(v_0) \cup S(v_1) \cup \cdots \cup S(v_{k-1})$ and $w \in S(v_k)$. Since each node is attached independently to w with the same probability p_w, it follows that

$$\mathbb{P}[\Gamma_w > k] = \begin{cases} q_w^{k+1} & k = 0, \ldots, n-1 \\ q_w^n & k \geq n. \end{cases}$$

This distribution does not depend on the order $v_0, v_1, v_2, \ldots, v_{n-1}$. For $t \geq 0$, we have

$$\phi_t = \prod_{\alpha \in S_{[t]}} q_\alpha = \prod_{j=0}^{t} \prod_{\alpha \in s(v_j) \setminus S[j-1]} q_\alpha \stackrel{d}{=} \prod_{j=0}^{t} \prod_{w \in W} q_w^{\mathbb{I}(\Gamma_w = j)} = \prod_{w \in W} q_w^{\mathbb{I}(\Gamma_w \leq t)}, \tag{20}$$

Using the fact that for a $B \sim \text{Bernoulli}(r)$, the expectation $\mathbb{E}[a^B] = 1 - (1-a)r$, we can easily calculate the expectation of ϕ_t

$$\begin{aligned} \mathbb{E}[\phi_t] &= \mathbb{E}[\prod_{w \in W} q_w^{\mathbb{I}(\Gamma_w \leq t)}] = \prod_{w \in W} \Big(1 - (1-q_w)\mathbb{P}[\Gamma_w \leq t)]\Big) \\ &= \prod_{w \in W} \Big(1 - (1-q_w)(1-q_w^{t+1})\Big). \end{aligned} \tag{21}$$

The concentration of ϕ_0 will be or crucial for the analysis of the supercritical regime, Subsection 5.2. Hence, we here provide $\mathbb{E}[\phi_0]$ and $\mathbb{E}[\phi_0^2]$. From (21) it follows

$$\mathbb{E}[\phi_0] = \prod_{w \in W}(1-p_w^2) = 1 - \sum_{w \in W} p_w^2 + o(\sum_{w \in W} p_w^2). \tag{22}$$

Moreover, from (20) it follows

$$\begin{aligned} \mathbb{E}[\phi_0^2] &= \mathbb{E}[\prod_{w \in W} q_w^{2\mathbb{I}(\Gamma_w \leq 0)}] = \prod_{w \in W} \Big(1 - (1-q_w^2)\mathbb{P}[\Gamma_w = 0)]\Big) = \prod_{w \in W} \Big(1 - (1-q_w^2)p_w\Big) \\ &= \prod_{w \in W} \Big(1 - 2p_w^2 + p_w^3\Big) = 1 - 2\sum_{w \in W} p_w^2 + \sum_{w \in W} p_w^3 + o(\sum_{w \in W} 2p_w^2 - p_w^3). \end{aligned} \tag{23}$$

5 Giant component

With the process (Y_t, N_t) defined in the previous section, we analyze both the subcritical and supercritical regime of our random intersection graph by adapting the percolation based techniques to analyze Erdős-Rényi random graphs [2].

The technical difficulty in analyzing that stopping time rests in the fact the distribution of Y_t depends on the history of the process. In the next two subsections, we will give conditions on $\{p_w : w \in W\}$ ensuring that $\mathbb{P}[|C(v)| \geq K \log n] < n^{-(1+\epsilon)}$ (subcritical regime) and $\mathbb{P}[|C(v)| < Kn] < XXXX$ (supercritical regime).

5.1 Subcritical regime

Theorem 5 *Let*

$$\sum_{w\in W} p_w^3 = O(1/n^2) \quad \textit{and} \quad p_w = O(1/n) \textit{ for all } w.$$

For any positive constant $c < 1$, if $\sum_{w\in W} p_w^2 \le c/n$, then all components in a random intersection graph are of order $O(\log n)$, with high probability.

Proof We generalize the techniques used in the proof for the sub-critical case in $G_{n,p}$ presented in [2]. Let $T(v_0)$ be the stopping time define in (11), for the process starting at node v_0 and note that $T(v_0) = |C(v_0)|$. We will bound the size of the largest component, and prove that under the conditions of the theorem, all components are of order $O(\log n)$, with high probability.

For all $t \ge 0$,

$$\mathbb{P}[T(v_0) > t] \le \mathbb{P}[Y_t > 0] = \mathbb{P}[\mathrm{Bin}(n-1, 1 - \prod_{\tau=0}^{t-1}(1-r_\tau)) \ge t]. \tag{24}$$

Bounding from above, which can easily be proven by induction in t for $r_\tau \in [0,1]$, we have

$$1 - \prod_{\tau=0}^{t-1}(1-r_\tau) \le \sum_{\tau=0}^{t-1} r_\tau = \sum_{\tau=0}^{t-1}(\phi_{\tau-1} - \phi_\tau) = 1 - \phi_{t-1}. \tag{25}$$

Further, by using stochastic ordering of the Binomial distribution (both in n and in $\sum r_\tau$), it follows

$$\begin{aligned}
\mathbb{P}[T > t] &\le \mathbb{P}[\mathrm{Bin}(n, \sum_{\tau=0}^{t-1} r_\tau) \ge t] = \mathbb{P}[\mathrm{Bin}(n, 1-\phi_{t-1}) \ge t] \\
&= \mathbb{P}[\mathrm{Bin}(n, 1-\phi_{t-1}) \ge t \mid 1-\phi_{t-1} < t/n]\mathbb{P}[1-\phi_{t-1} < t/n] \\
&\quad + \mathbb{P}[\mathrm{Bin}(n, 1-\phi_{t-1}) \ge t \mid 1-\phi_{t-1} > t/n]\mathbb{P}[1-\phi_{t-1} > t/n] \\
&\le \mathbb{P}[\mathrm{Bin}(n, 1-\phi_{t-1}) \ge t \mid 1-\phi_{t-1} < t/n] + \mathbb{P}[1-\phi_{t-1} \ge t/n].
\end{aligned} \tag{26}$$

For $t = K_0 \log n$, where K_0 is large enough and independent on the initial node v_0, the Chernoff bound ensures that $\mathbb{P}[\mathrm{Bin}(n, 1-\phi_{t-1}) \ge t \mid 1-\phi_{t-1} < t/n] = o(n^{-1})$. To bound $\mathbb{P}[1-\phi_{t-1} \ge t/n]$, use (20) to obtain

$$\begin{aligned}
\{1-\phi_{t-1} \ge t/n\} &= \left\{\prod_{w\in W} q_w^{\mathbb{I}(\Gamma_w \le t)} \le 1 - \frac{t}{n}\right\} \\
&= \left\{\sum_{w\in W} \log\left(\frac{1}{1-p_w}\right)\mathbb{I}(\Gamma_w \le t) \ge -\log\left(1-\frac{t}{n}\right)\right\}.
\end{aligned}$$

Linearize $-\log(1-t/n) = t/n + o(t/n)$ and define the bounded auxiliary random variables $X_{t,w} = n\log(1/(1-p_w))\mathbb{I}(\Gamma_w \le t)$. Direct calculations reveal that

$$\begin{aligned}
\mathbb{E}[X_{t,w}] &= n\log\Big(\frac{1}{1-p_w}\Big)(1-q_w^t) = n\Big(p_w + o(p_w)\Big)\Big(1-(1-p_w)^t\Big) \\
&= n\Big(p_w + o(p_w)\Big)\Big(tp_w + o(tp_w))\Big) = ntp_w^2 + o\Big(ntp_w^2\Big),
\end{aligned} \tag{27}$$

which implies

$$\sum_{w\in W} \mathbb{E}[X_{t,w}] = nt\sum_{w\in W} p_w^2 + o\Big(nt\sum_{w\in W} p_w^2\Big). \tag{28}$$

Thus under the stated condition that

$$n\sum_{w\in W} p_w^2 \le c < 1,$$

it follows that $0 < (1-c)t \le t - \sum_{w\in W} \mathbb{E}[X_{t,w}]$. In light of Bernstein's inequality, we bound

$$\begin{aligned} \mathbb{P}[1-\phi_{t-1} \ge t/n] &= \mathbb{P}\left[\sum_{w\in W} X_{t,w} \ge t\right] \le \mathbb{P}\left[\sum_{w\in W} X_{t,w} - \mathbb{E}[X_{t,w}] \ge (1-c)t\right] \\ \le \quad & \exp\Big(-\frac{\frac{3}{2}((1-c)t)^2}{3\sum_{w\in W}\mathrm{Var}[X_{t,w}] + nt\max_w\{p_w\}(1+o(1))}\Big). \end{aligned} \tag{29}$$

Since

$$\begin{aligned} \mathbb{E}[X_{t,w}^2] &= \Big(n\log\Big(\frac{1}{1-p_w}\Big)\Big)^2(1-q_w^t) = n^2\Big(p_w + o(p_w)\Big)^2\Big(1-(1-p_w)^t\Big) \\ &= n^2\Big(p_w^2 + o(p_w^2)\Big)\Big(tp_w + o(tp_w)\Big)\Big) = n^2tp_w^3 + o\Big(n^2t\sum_{w\in W} p_w^3\Big), \end{aligned} \tag{30}$$

it follows that for some large $K_1 > 0$

$$\sum_{w\in W}\mathrm{Var}[X_{t,w}] \le \sum_{w\in W}\mathbb{E}[X_{t,w}^2] = n^2t\sum_{w\in W}p_w^3 + o\Big(n^2t\sum_{w\in W}p_w^3\Big) \le K_1 t.$$

Finally, the assumption of the theorem implies that there exists $K_2 > 0$ such that

$$n\max_{w\in W} p_w \le K_2.$$

Substituting these bounds into (29) yields

$$\mathbb{P}[1-\phi_{t-1} \ge t/n] \le \exp\left(-\frac{3(1-c)^2}{2(3K_1+K_2)}t\right),$$

and taking $t = K_3\log n$ for some K_3 large enough and not depending on the initial node v_0, we conclude that $\mathbb{P}[1-\phi_{t-1} \ge t/n] = o(n^{-1})$, which in turn implies that taking $K_4 = \max\{K_0, K_3\}$, ensures that

$$\mathbb{P}[T(v_0) > K_4\log n] = o(1/n)$$

for any initial node v_0. Finally, a union bound over the n possible starting values v_0 implies that

$$\mathbb{P}[\max_{v_0\in V} T(v_0) > K_4\log n] \le no(n^{-1}) = o(1),$$

which implies that all connected components in the random intersection are of size $O(\log n)$, with probability tending to one as n tends to infinity. ■

Let us now comment that the conditions of the theorem are not self-redundant. From Cauchy-Schwarz inequality it follows that $\Big(\sum_{w\in W} p_w^3\Big)\Big(\sum_{w\in W} p_w\Big) \ge \Big(\sum_{w\in W} p_w^2\Big)^2 = c^2/n^2$. Given that $\sum_{w\in W} p_w^3 = O(1/n^2)$ it follows that $\sum_{w\in W} p_w = \Omega(1)$. From the fact that $p_w = O(1/n)$, it follows that $\sum_{w\in W} p_w = O(m/n) = \Omega(1)$.Thus this is valid case only if $m = \Omega(n)$.

5.2 Supercritical regime

In this subsection we consider the (very) supercritical regime, where $\sum_{w\in W} p_w^2 = c/n$, where $c > 1$ is a constant. The barely supercritical regime, when $c > 1$ but not a constant, we leave for future work.

Theorem 6 *Let*

$$\sum_{w\in W} p_w^3 = O(\frac{1}{n\log n}) \quad and \quad p_w = O(1/\log n) \, for\ all\ w.$$

For any positive constant $c > 1$, if $\sum_{w\in W} p_w^2 = c/n$, then there is a unique largest component of order $\Theta(n)$, in a random intersection graph, with high probability. Moreover, the size of the largest component is given by $(1+o(1))(1-\rho)$, where ρ is the solution in $(0,1)$ of the equation $\rho = \exp(c(\rho-1))$. All other components are of size $O(\log n)$.

First, note that the conditions imposed on p_w are weaker than ones in the case of the sub-critical regime.

The outline of our proof is the following. We first consider the value $1 - \prod_{\tau=0}^{t-1}(1-r_\tau)$, which is of essential for a behavior of the branching processes (Y_t, N_t), see Lemma 3. We provide lower and upper bounds for that value. We first, show that the process can be stochastically lower bounded by $\mathrm{Bin}(n, c'/n)$, for some constant $c' > 1$. This will ensure the stopping time in RIG for some node v_0 is of size $\Theta(n)$, **whp** [2]. Hence, there will be a component of size $\Theta(n)$ in our RIG. Furthermore, we provide an upper bound for the value $1 - \prod_{\tau=0}^{t-1}(1-r_\tau)$, showing the upper stochastic dominance. Given those two stochastic bounds, from the analysis on the corresponding Erdős-Rényi graphs, we conclude that the sizes of all components in RIG are of order either $O(\log n)$ or $\Theta(n)$. This of course does not provide that there is a unique largest component in RIG, since the proof for the uniqueness in Erdős-Renyi graphs depends on its structure. We finally show the uniqueness of the largest component and show that its size follows the law similar for the Erdős-Rényi graphs with the corresponding characteristics.

Proof

Let us consider the value $1 - \prod_{\tau=0}^{t-1}(1-r_\tau)$. One upper bound, $\sum_{\tau=0}^{t-1} r_\tau$, is already given in (25). Using Jensen's inequality on the function $\log(1-x)$ we provide a lower bound

$$\begin{aligned}\log\prod_{\tau=0}^{t-1}(1-r_\tau) &= \sum_{\tau=0}^{t-1}\log(1-r_\tau) = \sum_{\tau=0}^{t-1}\log\Big(1-(\phi_{\tau-1}-\phi_\tau)\Big) \\ &\le t\log\Big(1-\frac{1}{t}\sum_{\tau=0}^{t-1}(\phi_{\tau-1}-\phi_\tau)\Big) = t\log\Big(1-\frac{1-\phi_{t-1}}{t}\Big). \qquad (31)\end{aligned}$$

Thus, the value $1 - \prod_{\tau=0}^{t-1}(1-r_\tau)$ is lower and upper bounded by

$$1-\Big(1-\frac{1-\phi_{t-1}}{t}\Big)^t \le 1-\prod_{\tau=0}^{t-1}(1-r_\tau) \le \sum_{\tau=0}^{t-1} r_\tau = 1-\phi_{t-1}. \qquad (32)$$

In order to lower bound $1-\Big(1-\frac{1-\phi_{t-1}}{t}\Big)^t$, let us introduce and analyze $f_t(x) = 1-(1-x/t)^t$, where $t \ge 1$. It will be of our interest to consider $f_t(x)$ when $x \downarrow 0$. First the $f_t(x)$ is decreasing in t for fixed x. Hence $f_t(x) \ge 1-e^{-x}$. Moreover, for any $\epsilon > 0$ (ϵ is not necessarily a constant, and will be explained later) we have that for $x \in [0,\epsilon]$, we have $\frac{1-e^{-\epsilon}}{\epsilon}x \le 1-e^{-x} \le f_t(x)$. Second $f_t(x)$ is increasing in x for fixed t. From (20), $1-\phi_0 \le 1-\phi_t$, hence $1-(1-\frac{1-\phi_0}{t})^t \le 1-(1-\frac{1-\phi_{t-1}}{t})^t$. Now, let us look at $1-\phi_0$. From (23) and (22), by using Chebyshev inequality, with $\sum_{w\in W} p_w^2 = c/n$, it follows that ϕ_0, is concentrated around its mean, that is, for any constant $\delta > 0$, $\phi_0 \in (1\pm\delta)\mathbb{E}[\phi_0]$, with probability $1-o(1/n)$. Thus, it follows that, for any constant $\delta > 0$, $1-\phi_0 \in (1\pm\delta)c/n$, with probability $1-o(1/n)$. Now, we conclude for any $\delta > 0$ there is $\epsilon > 0$ such that $(c-\delta)\frac{1-e^{-\epsilon}}{\epsilon} > 1$, since constant $c > 1$. Note that $\lim_{\epsilon\to 0}\frac{1-e^{-\epsilon}}{\epsilon} = 1$, thus by choosing ϵ sufficiently small, then $\frac{1-e^{-\epsilon}}{\epsilon}$ can be arbitrarily close to 1. Hence, we have that $1-\prod_{\tau=0}^{t-1}(1-r_\tau) > c'/n$, for some constant $c^- > 1$, which is arbitrarily close to c. This means that our branching process on RIG is stochastically lower bounded by the $\mathrm{Bin}(n-1, c'/n)$, which stochastically dominates a branching process on $G_{n,c'/n}$. We know that there is a

[2] We will denote **whp**, meaning with the probability tending to 1, as the number of nodes n tends to infinity.

giant component in $G_{n,c'/n}$, of size $\Theta(n)$, that is, its stopping time is of order $\Theta(n)$ and so is the stopping time for RIG. Hence there is a giant component in RIG, **whp**.

The next paragraphs bound the size of the giant component in RIG. Thus we consider the regime for which exists a giant component, that is, starting from some v_0, its stopping time $T_{v_0} = \Theta(n)$.

We now show, that for the conditions of the theorem there is $p^+ = c^+/n$, for some constant $c^+ > 0$, such that $1 - \phi_{t-1} \leq 1 - (1-p^+)^t$. This is equivalent to $-\log \phi_{t-1} \leq t \log(1-p^+) = tp^+ + o(tp^+) = tc^+/n + o(t/n)$. From the definition for ϕ_{t-1}, see (20), with previously introduced random variables $X_{t,w} = n \log(1/(1 - p_w))\mathbb{I}(\Gamma_w \leq t)$, it follows that we want to show that $\mathbb{P}[\sum_w X_{t,w} < tc^+]$. In the previous subsection, we have analyzed $X_w, \mathbb{E}[X], \mathbb{E}[X_w], \mathbb{E}[X_w^2]$. Given that $\sum_w \mathbb{E}[X_{t,w}] = c/n$, by choosing $c^+ > c > 1$, by Bernstein inequality it follows that $\mathbb{P}[\sum_w X_{t,w} < tc^+] = 1 - o(1/n)$, for the conditions of the theorem.

Analogously, we now show, that for the conditions of the theorem there is $p^- = c^-/n$, for some constant $c^- > 1$, such that $1 - (1-p^-)^t \leq 1 - (1 - (1-\phi_{t-1})/t)^t$. This is equivalent to $-\log \phi_{t-1} \geq t \log(1-p^-) = tp^- + o(tp^-) = tc^-/n + o(t/n)$. With the analogue conclusion as in the previous case, by choosing $c > c^- > 1$ by Bernstein inequality it follows that $\mathbb{P}[\sum_w X_{t,w} > tc^-] = 1 - o(1/n)$, for the conditions of the theorem.

Thus, with probability is $1 - o(1/n)$, we have

$$1 - (1-p^-)^t \leq 1 - \prod_{\tau=0}^{t-1}(1 - r_\tau) \leq 1 - (1-p^+)^t. \tag{33}$$

From the stochastic dominance (33), we have

$$\begin{aligned} & \mathbb{P}\Big[\mathrm{Bin}\Big(n-1, 1-(1-p^-)^t\Big) \geq t\Big] \leq \mathbb{P}\Big[\mathrm{Bin}\Big(n-1, 1-\prod_{\tau=0}^{t-1}(1-r_\tau)\Big) \geq t\Big] \\ \leq \; & \mathbb{P}\Big[\mathrm{Bin}\Big(n-1, 1-(1-p^+)^t\Big) \geq t\Big]. \end{aligned} \tag{34}$$

Bellow we argue the uniqueness. RIG cannot be interpreted as a $G_{n,p}$, with a certain edge probability, since the edges in RIG are not independent, see (2). From (33) it follows that if we consider the set the recurrent equations (Y_t, N_t) on RIG, is stochastically bounded by the sets of the corresponding equations on G_{n,p^-} and G_{n,p^+}. Since c^-, c^+ are such that there are giant components in G_{n,p^-} and G_{n,p^+} **whp**, it follows that there must exist a giant component in our RIG, **whp**.

We now look closer at the size of a giant component. For the case when $p = \lambda/n$, where $\lambda > 1$, in [Hofstad, Spencer] it has been shown that there is the unique connected component in $G_{n,p}$, of size $\approx n\zeta_\lambda$, where ζ_λ is the unique solution from $(0, 1)$ of the equation

$$1 - e^{-\lambda\zeta} = \lambda. \tag{35}$$

More precisely, in [Hofstad], the CLT for the size of the giant component in $G_{n,\lambda/n}$ has been proven, which in distribution follows

$$\frac{|C_{\max}| - \zeta_\lambda n}{\sqrt{n}} \to \mathcal{N}\Big(0, \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda+\lambda\zeta_\lambda)^2}\Big), \tag{36}$$

where $|C_{\max}| = \max_v\{|C(v)|\}$. Remember that the stopping time $T = \inf\{t \geq 0 : Y_t = 0\}$, see (24). Hence, (34) and (36) imply that in random intersection graphs, **whp** there is a giant component of size, at least, $n\zeta_\lambda(1 - o(1))$. Considering, G_{n,p^-}, G_{n,p^+}, we know that their stopping times will be approximately ζn, where ζ satisfy (35), for $\lambda^- = np^-$, $\lambda^+ = np^+$, respectively. More precisely the giant component in G_{n,p^-}, G_{n,p^+} will follow (36). From (34) the stopping time of RIG is bounded by the stopping times on these two random graphs. To argue that those two times are closed, let us look at $F(\zeta, c) = 1 - \zeta - e^{-c\zeta}$, where (ζ, c) be the solution of $F(\zeta, c) = 0$, for given c. Since all partial derivatives of $F(\zeta, c)$ are continuous and bounded around (c, ζ), it follows that for a small perturbation of c, the solution in ζ cannot deviate much.

Now, our goal is to show that **whp** there is a unique giant component for the conditions of the theorem. We pursue the proof on RIG, analogously by adopting the proof for the uniqueness of the giant component in $G_{n,p}$, from [Joel-Spencer-Notes]. Let us assume that there are at least two giant components in RIG, denoted $V_1, V_2 \subset V$. Let us consider a new independent "sprinkling" RIG$'$ on the 'top' of our RIG, that is consider RIG$^+$ = RIG $\cup$ RIG$'$. Let for any $w \in W$ be $p'_w = p_w^\gamma$, where $\gamma > 1$ to be defined later. Let us consider all $\Theta(n^2)$ pairs $\{v_1, v_2\}$, where $v_1 \in V_1, v_2 \in V_2$, which are independent in RIG$'$ but not in RIG. Hence, the probability that two nodes $v_1, v_2 \in V$ are connected in RIG$'$ is given by

$$1 - \prod_w (1 - p'^2_w) = 1 - \prod_w (1 - p_w^{2\gamma}) = \sum_w p_w^{2\gamma} + o(\sum_w p_w^{2\gamma}) = \omega(1/n^2). \tag{37}$$

The last equality is true, since f$\gamma > 1$ and $p_w = O(1/n)$ for any w. By Markov inequality, it follows that there is a pair $\{v_1, v_2\}$, such that v_1 is connected to v_2 in RIG$'$. Thus, the components V_1, V_2 are connected, **whp**, forming one connected component within RIG$^+$. Then it follows that this component is of size at least $2n\zeta_\lambda(1-\delta)$, for some sufficiently small constant $\delta > 0$. On the other hand, since the probabilities in RIG$^+$ are

$$p_w^+ = 1 - (1 - p_w)(1 - p'_w) = p_w + p'_w(1 - p_w) = p_w + p_w^\gamma(1 - p_w) = p_w(1 + o(1)),$$

again this is true, since $\gamma > 1$ and $p_w = O(1/n)$ for any w. Thus,

$$\sum_{w \in W} (p_w^+)^2 = \sum_{w \in W} p_w^2 + \Theta(\sum_{w \in W} p_w^{1+\gamma}(1 - p_w)) = \sum_{w \in W} p_w^2(1 + o(1)) = c/n + o(1/n). \tag{38}$$

From (36), bounds on the stopping time in RIG and its continuity, it follows that the giant component in RIG$^+$ cannot be that large. This is a contradiction, thus there is only one giant component in RIG, of size given by $n\zeta(1 \pm o(1))$, where ζ satisfies (35). Moreover from (34) it follows that all other components are of size $O(\log n)$. ∎

References

[1] ALBERT, R., AND BARABÁSI, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys. 74*, 1 (2002), 47 – 97.

[2] ALON, N., AND SPENCER, J. H. *The probabilistic method*, 2nd ed. John Wiley & Sons, Inc., New York, 2000.

[3] BARABÁSI, A. L., AND ALBERT, R. Emergence of Scaling in Random Networks. *Science 286*, 5439 (1999), 509–512.

[4] BEHRISCH, M. Component evolution in random intersection graphs. In *Electr. J. Comb.* (2007), vol. 14.

[5] BLOZNELIS, M., JAWORSKI, J., AND RYBARCZYK, K. Component evolution in a secure wireless sensor network. *Netw. 53*, 1 (2009), 19–26.

[6] BROMAN, E., AND MEESTER, R. Survival of inhomogeneous galton-watson processes. *Adv. in Appl. Probab. 40*, 3 (2008), 798–814.

[7] CHUNG, F., AND LU, L. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America 99*, 25 (2002), 15879–15882.

[8] DEIJFEN, M., AND KETS, W. Random intersection graphs with tunable degree distribution and clustering. *Probab. Eng. Inf. Sci. 23*, 4 (2009), 661–674.

[9] ERDŐS, P., GOODMAN, A. W., AND PÓSA, L. The representation of a graph by set intersections. *Canad. J. Math. 18* (1966), 106–112.

[10] ERHARD GODEHARDT, JERZY JAWORSKI, K. R. Random intersection graphs and classification. In *Advances in Data Analysis* (2007), vol. 45, pp. 67–74.

[11] EUBANK, S., GUCLU, H., ANIL KUMAR, V. S., MARATHE, M. V., SRINIVASAN, A., TOROCZKAI, Z., AND WANG, N. Modelling disease outbreaks in realistic urban social networks. *Nature 429*, 6988 (May 2004), 180–184.

[12] GODEHARDT, E., AND JAWORSKI, J. Two models of random intersection graphs and their applications. *Electronic Notes in Discrete Mathematics 10* (2001), 129–132.

[13] GUILLAUME, J.-L., AND LATAPY, M. Bipartite graphs as models of complex networks. *Physica A: Statistical and Theoretical Physics 371*, 2 (2006), 795 – 813.

[14] KAROŃSKI, M., SCHEINERMAN, E., AND SINGER-COHEN, K. On random intersection graphs:the subgraph problem. *Combinatorics, Probability and Computing 8* (1999).

[15] NEWMAN, M. E. J. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E 64*, 1 (Jun 2001), 016131.

[16] NEWMAN, M. E. J., AND PARK, J. Why social networks are different from other types of networks. *Phys. Rev. E 68*, 3 (Sep 2003), 036122.

[17] NEWMAN, M. E. J., STROGATZ, S. H., AND WATTS, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E 64*, 2 (Jul 2001), 026118.

[18] NIKOLETSEAS, S., RAPTOPOULOS, C., AND SPIRAKIS, P. Large independent sets in general random intersection graphs. *Theor. Comput. Sci. 406* (October 2008), 215–224.

[19] NIKOLETSEAS, S. E., RAPTOPOULOS, C., AND SPIRAKIS, P. G. The existence and efficient construction of large independent sets in general random intersection graphs. In *ICALP* (2004), J. Daz, J. Karhumki, A. Lepist, and D. Sannella, Eds., vol. 3142 of *Lecture Notes in Computer Science*, Springer, pp. 1029–1040.

[20] NIKOLETSEAS, S. E., RAPTOPOULOS, C., AND SPIRAKIS, P. G. Expander properties and the cover time of random intersection graphs. *Theor. Comput. Sci. 410*, 50 (2009), 5261–5272.

[21] REMCO VAN DER HOFSTAD. Random graphs and complex networks. Lecture notes in preparation, `http://www.win.tue.nl/~rhofstad/NotesRGCN.html`.

[22] SINGER-COHEN, K. Random intersection graphs. PhD thesis, Johns Hopkins University, 1995.

[23] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of Small-World networks. *Nature 393*, 6684 (1998), 440–442.