Component evolution in random intersection graphs

Michael Behrisch *

Institut für Informatik, Humboldt-Universität zu Berlin, 10099 Berlin, Germany
behrisch@informatik.hu-berlin.de

Submitted: Jan 21, 2005; Accepted: Oct 26, 2006; Published: Jan 29, 2007
Mathematics Subject Classification: 05C80

Abstract

We study the evolution of the order of the largest component in the random intersection graph model which reflects some clustering properties of real–world networks. We show that for appropriate choice of the parameters random intersection graphs differ from $G_{n,p}$ in that neither the so-called giant component, appearing when the expected vertex degree gets larger than one, has linear order nor is the second largest of logarithmic order. We also describe a test of our result on a protein similarity network.

1 Introduction

The classical random graph model (introduced by Erdős and Rényi in the early 1960s) considers a fixed set of n vertices and edges that exist with a certain probability $p = p(n)$, independently from each other. It was shown to be inappropriate for describing real–world networks because it lacks certain features of those (e.g. scale free degree distribution and clustering). One of the underlying reasons that are responsible for this mismatch is precisely the independence of the edges, in other words the missing transitivity. In a real–world network, relations between vertices x and y on the one hand and between vertices y and z on the other hand suggest a connection of some sort between vertices x and z.

An *intersection graph* is a graph on vertex set V where each vertex has a subset of a ground set W assigned and two vertices are adjacent if and only if the assigned sets have a non-empty intersection.

We call the ground set W from which the assigned sets are chosen *universal feature set* and its elements *features*. Furthermore the set of vertices V_w holding a specified feature w (which obviously forms a clique) is called *feature clique* while W_v shall denote the *feature set* assigned to vertex v. We generalise this notation to sets of vertices and features in the obvious way, e.g. $W_U = \bigcup_{u \in U} W_u$.

*supported by the DFG research center MATHEON in Berlin.

Examples for intersection graphs are the well studied interval graphs on the real line, in this paper however we will only consider finite sets.

A random intersection graph on n vertices with a universal feature set of size m is one where each vertex chooses each feature independently with probability p. A sample of this probability space is denoted by $G_{n,m,p}$.

We consider now and in the following at $m := n^\alpha$ with either $\alpha > 1$ or $0 < \alpha < 1$.

This random model was invented and studied with respect to subgraph appearance by Karoński, Scheinerman and Singer-Cohen in [11], with respect to equivalence to $G_{n,p}$ by Fill, Scheinerman, Singer-Cohen in [7] and with respect to vertex degree distribution by Stark [14]. An algorithmic reconstruction of the feature structure with only the intersection graph as input was given by Behrisch and Taraz in [2]. The first two results and some results concerning connectivity and cliques can also be found in Singer [13]. Recently also the chromatic and the independence number of random intersection graphs have been investigated by Behrisch, Taraz, and Ueckerdt [15, 3].

Extensions to the model were proposed by Godehardt and Jaworski in [8], who modify the distribution of the sizes of the feature cliques and practical relevance of the model was studied by Newman, Strogatz and Watts in [12] and by Guillaume and Latapy in [9].

The aim of this paper is to study the evolution of the largest component in this model. Since components are natural candidates for clusters in graphs it is straightforward to analyse their growth in our random model, thereby getting insight into structural peculiarities of the real–world networks. The component structure for $G_{n,p}$ was already studied by Erdős and Rényi in [6] and there are also results for some models for real–world networks by Chung and Lu [5] and Bollobás and Riordan [4].

The paper is organised as follows. In the next section we describe our results and compare it with the growth of the giant component in $G_{n,p}$. Section 3 states some results on branching processes which will be used for the proofs of the results in Section 4 and 5. We close with experimental studies on the evolution of a real–world network.

2 The results

Let $\mathcal{N}(G)$ denote the order (number of vertices) of the largest component of G. Our main theorem is:

Theorem 1. *Let $G_{n,m,p}$ be a random intersection graph with $m := n^\alpha$ and $p^2 m = \frac{c}{n}$. Furthermore let ρ be the single solution to $\rho = \exp(c(\rho - 1))$ in the interval $(0, 1)$ for $c > 1$ Then we have a.a.s.*

$$\mathcal{N}(G_{n,m,p}) \leq \frac{9}{(1-c)^2} \ln n \qquad \text{for } \alpha > 1 \text{ and } c < 1 \tag{1}$$

$$\mathcal{N}(G_{n,m,p}) = (1 + o(1))(1 - \rho)n \qquad \text{for } \alpha > 1 \text{ and } c > 1 \tag{2}$$

$$\mathcal{N}(G_{n,m,p}) \leq \frac{10\sqrt{c}}{(1-c)^2} \sqrt{\frac{n}{m}} \ln m \qquad \text{for } \alpha < 1 \text{ and } c < 1 \tag{3}$$

$$\mathcal{N}(G_{n,m,p}) = (1 + o(1))(1 - \rho)\sqrt{cmn} \qquad \text{for } \alpha < 1 \text{ and } c > 1 \tag{4}$$

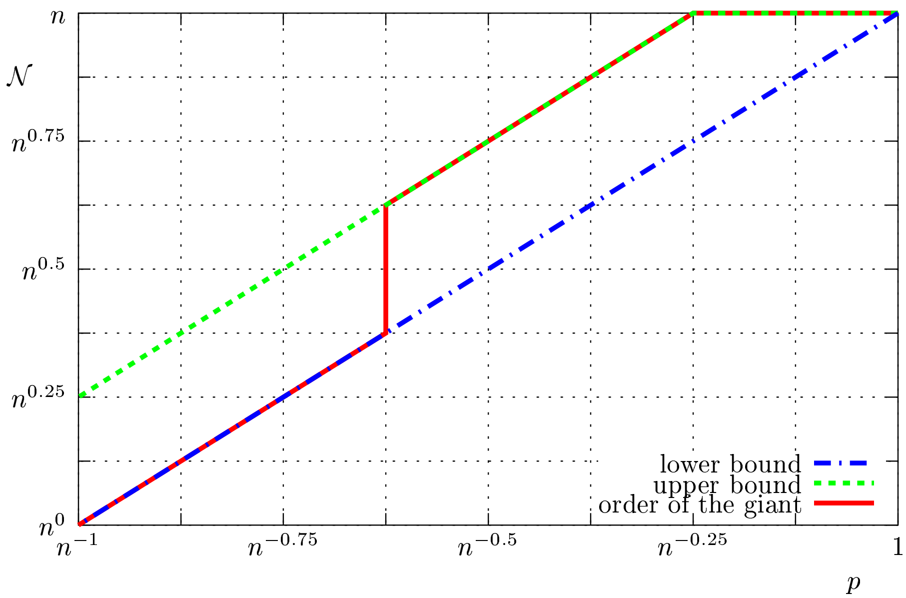

Figure 1: Evolution of the largest component for $\alpha = 0.25$.

Furthermore we can prove that the order of the largest component for $\alpha < \frac{1}{2}$ and p small enough is approximately that of a single feature clique, see Section 5.1 for details.

As already proven in [13] the "edge probability" p' (meaning the ratio between present edges and all possible edges) in the random intersection graph is closely concentrated around p^2m. Thus the two results above show that for $\alpha > 1$ the largest component in the intersection graph exhibits a jump from logarithmic order to linear order at $p' = \frac{1}{n}$ which is similar to the $G_{n,p'}$ behaviour. This is also the moment at which in both models the expected degree of a vertex gets larger than 1.

For $\alpha < 1$ the jump is still at the same position but $\mathcal{N}$ increases only by a polynomial factor as is shown in Figure 1 for $\alpha = 0.25$.

Additionally this figure shows that the order of the largest component jumps from approximately the size of a single feature clique (which is concentrated around pn, see (12)) as a trivial lower bound to the order of the largest component to approximately the sum of the sizes of all feature cliques (which is for the same reasons concentrated around pmn) which is an upper bound to $\mathcal{N}$.

3 Branching processes and auxiliary lemmas

In order to discover components in a graph we will use branching processes (for an overview of the topic of branching processes and for references to proofs see [1]) similar to the proofs in Chapter 5 of [10]. We will explore the component by starting at a single vertex, generating its neighbors as descendants in a branching process and then the second neighbourhood as their descendants and so forth.

Let the random variable X with binomial distribution $\mathrm{Bi}(n,p)$ denote the number of descendants (neighbors) of an arbitrary vertex. The Galton-Watson branching process on the variable X has the following properties (see Theorem 5.1 and Example 5.2 and 5.3 in [10]).

1. If $np \xrightarrow{n\to\infty} c < 1$ the branching process on X dies out a.a.s.
2. If $np \xrightarrow{n\to\infty} c > 1$ the branching process dies out with probability $\rho(c)$ where $\rho(c)$ is the unique solution of
$$\rho = \exp(c(\rho - 1)) \tag{5}$$
in the interval $(0,1)$.

Thus the main complication in the proof is to overcome the limitations of the branching process which deals with an essentially unbounded domain in contrast to the limited number of vertices in the graph.

The discovery of neighbors is (in contrast to the process used in the $G_{n,p}$ model) a two step process. First we let the vertex discover its features and then the features find the vertices they are assigned to. The features and the vertices used in each step will be ignored in the further process which will slightly downsize the universal feature set and the vertex set. As we will see later this deviation will not affect the ongoing process very much.

3.1 Auxiliary lemmas

The following estimates are used without proof:

$$(1-a)^b = (1+o(1))(1-ab) \qquad \text{for } 0 < a < 1, ab \to 0 \tag{6}$$

$$\mathrm{e}^{-2a} \le 1 - a \le \mathrm{e}^{-a} \qquad \text{for } 0 \le a \le \frac{1}{2} \tag{7}$$

Let X be a non-negative random variable with expectation $\mu := \mathbb{E}[X]$ and variance $\mathrm{Var}[X]$. As a special case of Markov's inequality the first moment method states that

$$\mathbb{P}[X \ge 1] \le \mu. \tag{8}$$

and the second moment method (special case of Tschebyscheff's inequality) that

$$\mathbb{P}[X = 0] \le \mathrm{Var}[X]/\mu^2 = \frac{\mathbb{E}[X^2]}{\mu^2} - 1. \tag{9}$$

If X is a binomially distributed random variable (n trials, each with probability p), then $\mu = np$ and we shall use the following variants of Chernoff's inequality (see Section 2 in [10]):

$$\mathbb{P}[X \ge \mu + t] \le \exp\left(-\frac{t^2}{2(\mu + t/3)}\right) \qquad \text{for } t \ge 0, \tag{10}$$

$$\mathbb{P}[X \le \mu - t] \le \exp\left(-\frac{t^2}{2\mu}\right) \qquad \text{for } t \ge 0, \tag{11}$$

4 The evolution for $\alpha > 1$

This section contains the proof of the first two statements of Theorem 1. After giving a sharp concentration result on the number of features a single vertex may have, we closely resemble the branching process method used in [10] to prove the results on the order of the largest component.

4.1 The size of the feature set

In order to give precise estimates on the number vertices which get discovered by the branching process we need sharp bounds on the size of the feature set of a vertex.

Lemma 2. *Let v be a fixed vertex in a random intersection graph $G_{n,m,p}$ with $pn = o(1)$ and $p^2mn = \Theta(1)$. Furthermore let $W' \subseteq W$ be a subset of the universal feature set of size at least $m - 2pmn$ and $X_v := |W_v \cap W'|$ denote the random variable counting the number of features of v in W'. Then X_v is very likely close to its expectation or precisely:*

$$\mathbb{P}\left[|X_v - pm| > (pm)^{\frac{3}{4}}\right] \leq \exp\left(-\frac{(pm)^{\frac{1}{2}}}{3}\right)$$

Proof. For the expected number of features selected in W' we have $\mu := \mathbb{E}[X_v] \geq p(m - 2pmn)) = pm - O(1)$ and $\mu \leq pm$.

Since the features are selected independently uniformly at random we can use Chernoff inequalities (10) and (11) to bound the deviation from the expected size.

$$\begin{aligned}
\mathbb{P}\left[Y \geq pm + (pm)^{\frac{3}{4}}\right] &\leq \mathbb{P}\left[Y \geq \mu + (pm)^{\frac{3}{4}}\right] \\
&\leq \exp\left(-\frac{(pm)^{\frac{3}{2}}}{2\left(\mu + (pm)^{\frac{3}{4}}/3\right)}\right) \\
&\leq \exp\left(-\frac{(pm)^{\frac{3}{2}}}{2\left(pm + (pm)^{\frac{3}{4}}/3\right)}\right) \\
&\leq \frac{1}{2}\exp\left(-\frac{(pm)^{\frac{1}{2}}}{3}\right)
\end{aligned}$$

And for the lower tail using (11):

$$\begin{aligned}
\mathbb{P}\left[Y \leq pm - (pm)^{\frac{3}{4}}\right] &= \mathbb{P}\left[Y \geq \mu + O(1) - (pm)^{\frac{3}{4}}\right] \\
&\leq \exp\left(-\frac{\left((pm)^{\frac{3}{4}} - O(1)\right)^2}{2(pm - O(1))}\right) \\
&\leq \frac{1}{2}\exp\left(-\frac{(pm)^{\frac{1}{2}}}{3}\right)
\end{aligned}$$

Notice that these calculations (and thus the probability for the tails) remain valid even if we remove no features at all.

From the two tails above we may easily conclude the statement of the lemma. □

4.2 Proof of Theorem 1, (1) and (2)

Proof of (1). We prove that for $c < 1$ the branching process starting at an arbitrary vertex v discovering all the vertices one by one will finish in at most $\frac{9 \ln n}{(1-c)^2}$ steps.

From Lemma 2 we know that there is with high probability no large deviation from the expected value in the size of a feature set. Our branching process starting at v now proceeds as follows. At first v discovers its features. If there are too many or too few of them (in the sense of Lemma 2) we abort.

Otherwise we let the features discover the vertices which hold them. Since the feature set of v has size $(1 + o(1))pm$ the probability for an individual vertex w to hold at least one feature in this set is

$$\mathbb{P}\left[\{v, w\} \in E(G_{n,m,p})\right] = 1 - (1 - p)^{(1+o(1))pm} \overset{(6)}{=} (1 + o(1))p^2 m$$

and the neighbors of v will be chosen independently with this probability. Thus the expected number of new neighbors discovered will be:

$$\mathbb{E}\left[d(v)\right] \leq n(1 + o(1))p^2 m$$

Now we remove W_v (the feature set of v) from the universal feature set and continue with discovering the features of the neighbors of v the same way we discovered the features of v and so on. We do this at most n times (only n vertices available) thus the probability that we will abort at any step because of the wrong size of the feature set is (due to Lemma 2) bounded by

$$n \exp\left(-\frac{(pm)^{\frac{1}{2}}}{3}\right) \xrightarrow{n\to\infty} 0.$$

Furthermore we did remove at most $n(1 + o(1))pm < 2pmn$ features from the universal feature set thus Lemma 2 was applicable all the time.

Observe that the probability that v is in a component of order at least k is bounded by the probability that the sum of the degrees of k vertices discovered in the process is at least $k - 1$. Since all features were discovered independent from earlier ones and thus all vertices were discovered in an independent manner, the probability for a component of order at least $k \geq \frac{9 \ln n}{(1-c)^2}$ can be bounded using a Chernoff inequality again. Let Y_i denote the number of neighbors of the ith vertex discovered in the process and notice that the expected value for the sum over the Y_i is bounded from above by $(1 + o(1))kp^2 mn \leq kc'$ for $c' := \frac{c+1}{2}$.

$$\begin{aligned} n\mathbb{P}\left[\sum_{i=1}^{k} Y_i \geq k - 1\right] &= n\mathbb{P}\left[\sum_{i=1}^{k} Y_i \geq kc' + (1 - c')k - 1\right] \\ &\leq n \exp\left(-\frac{((1 - c')k - 1)^2}{2(c'k + (1 - c')k/3)}\right) \\ &\leq n \exp\left(-\frac{(1 - c')^2}{2} k\right). \end{aligned}$$

Resubstituting c' and k shows that this term tends to 0 as n tends to infinity which proves by (8) the theorem. $\square$

For the appearance of a giant component when $c > 1$ we will study the same branching process again using the proof of Janson, Łuczak and Rućinski [10].

Proof of (2). We start by proving that there is a.a.s. no component which has more than $k_- := \frac{50c}{(c-1)^2}\ln n$ or less than $k_+ := n^{2/3}$ vertices by proving the harder result that for every $k_< k < k_+$ there are a.a.s. $\frac{(c-1)}{2}k$ vertices which are to be examined (got discovered as neighbors but were not examined themselves). To prove this we have to look at no more than $k + \frac{c-1}{2}k = \frac{c+1}{2}k$ vertices.

Because of this we exclude in each step at most $\frac{c+1}{2}k_+$ vertices from the further process. Furthermore we do still downsize the universal feature set only for a very small amount for each vertex which discovers its neighbors as in the proof of (1). This gives independence for all steps of the branching process and thus one can bound the number of neighbors a vertex discovers from below by independent random variables $Y_i^* \in \mathrm{Bi}(n - \frac{c+1}{2}k_+, p'^2 m)$ with p' such that $p'^2 mn = \frac{3c+1}{4}$. The value for p' results from the lower bound on the size of feature set given by Lemma 2.

Now we can bound the probability of dying out after k steps or having too few discovered (but unexamined) vertices by the probability that

$$\sum_{i=1}^{k} Y_i^* \leq k - 1 + \frac{c-1}{2}k$$

Now the existence of such a process can be bound by Chernoff inequality (11) and we get with $\mu := \mathbb{E}\left[\sum_{i=1}^{k} Y_i^*\right] = \frac{3c+1}{4}k - o(k)$ for $k_- \leq k \leq k_+$ and n large enough:

$$\begin{aligned} n\sum_{k=k_-}^{k_+} \mathbb{P}\left[\sum_{i=1}^{k} Y_i^* \leq k - 1 + \frac{c-1}{2}k\right] &= n\sum_{k=k_-}^{k_+} \mathbb{P}\left[\sum_{i=1}^{k} Y_i^* \leq \mu - \left(\frac{c-1}{4}k - o(k) + 1\right)\right] \\ &\leq n\sum_{k=k_-}^{k_+} \exp\left(\frac{-\left(\frac{c-1}{4}k - o(k) + 1\right)^2}{\frac{3c+1}{2}k}\right) \\ &\leq n\sum_{k=k_-}^{k_+} \exp\left(\frac{-\left(\frac{c-1}{4}\right)^2 k}{3c}\right) \\ &\leq nk_+ \exp\left(\frac{-\left(\frac{c-1}{4}\right)^2 k_-}{3c}\right) \end{aligned}$$

Because of the values for k_- and k_+ given at the beginning of the proof this tends to 0 as n tends to infinity and thus by (8) there is a.a.s. no process stopping between k_- and k_+.

If there exist two different components T and U with $|T| \geq k_+$ and $|U| \geq k_+$ their sets of features W_T and W_U have to be disjoint. According to Lemma 2 a.a.s. $|W_U| \geq k_+\frac{pm}{2}$.

Thus the probability of disjointness is:

$$(1-p)^{k_+^2 \frac{pm}{2}} \overset{(7)}{\leq} \exp\left(-k_+^2 \frac{p^2 m}{2}\right) = \exp\left(-n^{\frac{4}{3}} \frac{c}{2n}\right) \xrightarrow{n\to\infty} 0$$

Now we have that there is a.a.s. only one component with at least k_+ vertices, it remains to show that it has linear order. Let Y, denote the number of vertices in components of order at most k_-. Let for each vertex $i \in V$ Y_i be the indicator variable for being in such a small component. We estimate expectation and variance of Y.

For a single vertex the probability of being in a small component can be bounded from above and from below by the extinction probabilities of branching processes with distribution $\mathrm{Bi}(n-k_-, (1-o(1))p^2 m)$ and $\mathrm{Bi}(n, (1+o(1))p^2 m)$. The $o(1)$ terms in the two cases bound the possible deviations in the size of feature sets according to Lemma 2. By (5) we know that the probability of extinction of these two processes is ρ which results by linearity of expectation into $\mathbb{E}[Y] = \rho(c)n$.

In order to prove the concentration of Y around its expectation, we calculate its variance, or precisely using (9) we show that $\mathbb{E}[Y^2] = (1+o(1))\mathbb{E}[Y]^2$. Two vertices being simultaneously in a small component is an event which occurs either if they are in the same component in that case the probability can be bounded by the extinction probability for this component or they are in two components which means two extinctions have to occur independently.

$$\begin{aligned}
\mathbb{E}\left[Y^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right] = \sum_{i,j} \mathbb{E}\left[Y_i Y_j\right] \\
&\leq n\rho(np)k_- + n\rho(np)n\rho((n-k_-)p) \\
&= (1+o(1))n^2\rho(np)^2 = (1+o(1))\mathbb{E}[Y]^2
\end{aligned}$$

By Tschebyscheff's inequality (9) we can conclude that the number of small vertices is a.a.s. $\rho(c)n$ hence the largest component is of order $(1-\rho(c))n$. □

One further consequence of this proof is that for $\alpha > 1$ and $c > 1$ we can bound the order of the second largest component by $\frac{50c}{(c-1)^2}\ln n$.

5 The evolution for $\alpha < 1$

If we have an small upper bound for the number of vertices two feature cliques have in common we can simply add the clique sizes (provided we know they are connected) in order to estimate the component order. This bound is the content of the following lemma.

Lemma 3. *Let Y be the random variable counting the number of vertices having more than one feature in a random intersection graph $G_{n,m,p}$ with $m := n^\alpha$ and $\alpha < 1$. Then for $p^2m^2n \gg \ln n$:*

$$\mathbb{P}\left[Y > 2p^2m^2n\right] \xrightarrow{n\to\infty} 0$$

and for $p^2m^2n \xrightarrow{n\to\infty} 0$*:*

$$\mathbb{P}\left[Y > 0\right] \xrightarrow{n\to\infty} 0$$

Proof. For a single fixed vertex v the probability of having more than one feature is (when $pm \to 0$):

$$\mathbb{P}\left[|W_v| > 1\right] = 1 - (1-p)^m - (mp(1-p)^{m-1} \overset{(6)}{=} (1+o(1))m^2p^2.$$

Since all vertices choose their features independently Y is a binomially distributed variable with expectation nm^2p^2 and the second statement of the lemma follows by Markov inequality. For the first statement we can bound the deviation using Chernoff inequality (10).

$$\mathbb{P}\left[Y > 2p^2m^2n\right] \leq \mathbb{P}\left[Y > 2\mathbb{E}\left[Y\right]\right] \leq \exp\left(-\frac{3nm^2p^2}{8}\right) \xrightarrow{n\to\infty} 0.$$

□

Now we can start proving the component evolution for $\alpha < 1$.

Proof of (3). In order to reuse the results of Section 4 we interchange the role of the feature set and the vertex set and look at the largest component in the feature set instead of one in the vertex set. As we know from Theorem (1) there will be no component containing more than $\frac{9}{(1-c)^2}\ln m$ features. Exploiting again the symmetry between feature set and vertex set, we can use Lemma 2 to deduce that for every feature w

$$V_w = (1+o(1))pn \tag{12}$$

with probability at least $1 - m\exp(-(pn)^{1/2}/3 = 1 - o(1))$. We can conclude that the order of the largest component is a.a.s. bounded by

$$\frac{9}{(1-c)^2}\ln m \cdot (1+o(1))pn \leq \frac{10\sqrt{c}}{(1-c)^2}\sqrt{\frac{n}{m}}\ln m.$$

□

Proof of (4). We use the same method as in the last proof. With exactly the same argument we already have a.a.s. an upper bound for the order of the largest component of

$$(1-\rho(c))m \cdot (1+o(1))pn \leq (1+o(1))\sqrt{c}(1-\rho(c))\sqrt{mn}.$$

The lower bound can be achieved because the order of the component can be bound by the sum over the sizes of all cliques minus the number of vertices which occur in more than one clique multiplied with the multiplicity they occur. Or more precise (with W_L denoting the set of features in the giant component in W and V_L denoting the vertices linked to it):

$$\begin{aligned}|V_L| &= \sum_{w\in W_L}|V_w| - \sum_{v\in V_L, |W_v|>1}(|W_v| - 1)\\ &\geq (1-\rho(c))m(1+o(1))pn - \sum_{v\in V_L, |W_v|>1}\max_{v\in V}\{|W_v|\}\end{aligned}$$

The probability for the existence of a vertex with more than $\ln m$ features is bounded by $n(pm)^{\ln m}$ which tends to 0 for our choice of p. Furthermore we know from Lemma 3 that there are at most $2p^2m^2n = 2cm$ vertices with more than one feature. Therefore

$$\begin{aligned}|V_L| &\geq (1-\rho(c))m(1+o(1))pn - 2cm\ln m\\ &= (1+o(1))(1-\rho(c))\sqrt{cmn} - 2cm\ln m\\ &= (1+o(1))(1-\rho(c))\sqrt{cmn}.\end{aligned}$$

□

As a direct consequence of this bound and the remark after the proof of (2) we have that for $\alpha < 1$ and $c > 1$ we can bound the order of the second largest component by $\frac{51c}{(c-1)^2}\ln mpn = \frac{51c\sqrt{c}}{(c-1)^2}\sqrt{\frac{n}{m}}\ln m$.

5.1 Feature cliques as components

Similar to the evolution of $G_{n,p}$, which has lots of isolated vertices for very small p, there are stages of the evolution of $G_{n,m,p}$ where the feature cliques do not intersect. The component structure of $G_{n,m,p}$ is very uncomplex then.

Proposition 4. *Let $G_{n,m,p}$ be a random intersection graph with $m := n^\alpha$ and $\alpha < \frac{1}{2}$ and $\ln n \ll pn \ll \frac{\sqrt{n}}{m}$. Then a.a.s. there are m components which are (feature) cliques and the rest of the graph consists of isolated vertices and thus a.a.s. $\mathcal{N}(G_{n,m,p}) = (1+o(1))pn$.*

Proof. The statement follows directly from Lemma 3 and (12) because if there are no vertices with more than one feature there are only isolated vertices and feature cliques. □

6 Experiments

We tested our result on an instance of a complete edge–weighted real world network on 5119 vertices. Here parts of proteins serve as vertices and the edge-weight describes their spatial similarity. If we look at the subgraph of this graph containing all edges with weight greater than a fixed value s (where greater edge weights indicate higher similarity) we can simulate an evolution of this network by gradually decreasing s. Thus first the highly analogue parts get connected and bit by bit also the less similar ones connect to the components.

The evolution found this way differs significantly from a graph in which the same weights are distributed uniformly at random among the edges (see Figure 2).

The most striking difference is the slow growth of the largest component in the stages after it has only very few vertices (minimum edge weight between 40 and 60). A similar behaviour cannot be modelled using standard random graphs where $\mathcal{N}$ is either logarithmic or linear in the number of vertices. As one can see in Figure 2 the random intersection graph resembles this steady aggregation of vertices to the largest component very well.

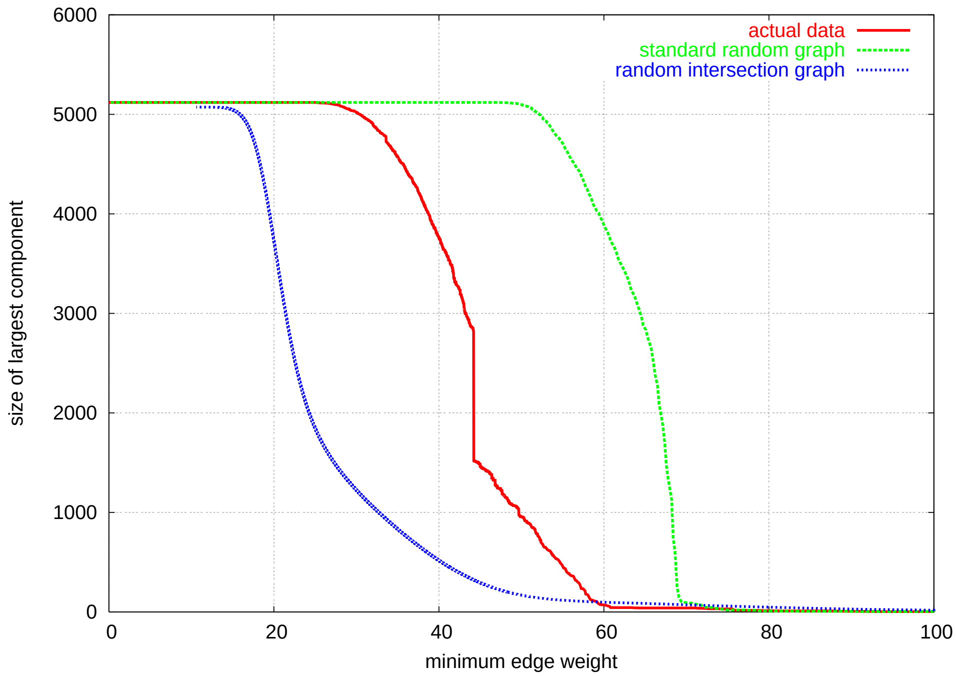

Figure 2: Evolution of the largest component in the protein graph.

References

[1] K. B. Athreya and A. N. Vidyashankar. Branching processes. Technical report, University of Georgia – Department of Statistics, 1999.

[2] M. Behrisch and A. Taraz. Efficiently covering complex networks with cliques of similar vertices. *Theoretical Computer Science*, 2005. to appear.

[3] M. Behrisch, A. Taraz, and M. Ueckerdt. Colouring random intersection graphs and complex networks. Preprint, December 2005.

[4] B. Bollobás and O. Riordan. Slow emergence of the giant component in the growing m-out graph. to appear in Random Structures and Algorithms.

[5] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145, 2002.

[6] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.

[7] J. A. Fill, E. R. Scheinerman, and K. B. Singer-Cohen. Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models. *Random Structures and Algorithms*, 16(2):156–176, March 2000.

[8] E. Godehardt and J. Jaworski. Two models of random intersection graphs and their applications. *Electronic Notes in Discrete Mathematics*, 10, 2001.

[9] J.-L. Guillaume and M. Latapy. Bipartite structure of all complex networks. *Information Processing Letters*, 90:215–221, 2004.

[10] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. John Wiley & Sons, 2000.

[11] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.

[12] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64, 2001.

[13] K. B. Singer. *Random Intersection Graphs*. PhD thesis, John Hopkins University, Baltimore, Maryland, 1995.

[14] D. Stark. The vertex degree distribution of random intersection graphs. *Random Structures and Algorithms*, 24(3):249–258, May 2004.

[15] M. Ueckerdt. Färben von zufälligen Schnittgraphen. Diploma thesis, Humboldt-Universität zu Berlin, 2005.