Two-point concentration in random geometric graphs

Tobias Müller
University of Oxford

27 March 2006

The Erdös-Rényi or $G(n, p)$-model

We construct a random graph with vertex set $V = \{1, \ldots, n\}$. For each candidates edge ij we flip a coin with success probability p to decide whether or not to include it, independently of all other edges.

Often one is interested in the behaviour of $G(n, p)$ as n grows large, where $p = p(n)$ is allowed to vary with n. Depending on the choice of $p(n)$ qualitatively different types of behaviour can be observed.

This model has been studied quite intensively over the past 5 decades and as a result quite a lot is known about it's behaviour which is often (surprisingly) complex.

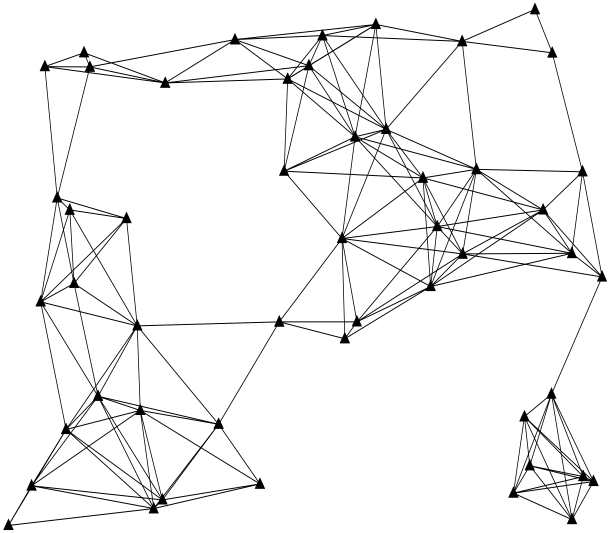
The random geometric model

We construct a random graph G_n as follows. We pick vertices $X_1, \ldots, X_n \in \mathbb{R}^d$ at random (iid according to some probability distribution ν on $\mathbb{R}^d$) and we join X_i, X_j ($i \neq j$) by an edge if $\| X_i - X_j \| < r$.

We are interested in the behaviour of the various graph parameters of this graph as n grows large, where $r = r(n)$ is allowed to vary with n. We will allow any choice of ν that has a bounded density function and $\| . \|$ may be any norm.

The distance $r(n)$ plays a role similar to that of $p(n)$ in $G(n, p)$. Depending on the choice of $r(n)$ qualitatively different types of behaviour can be observed.

An instance



Graph theory notation and terminology

Let $G = (V, E)$ be a graph.

$\Delta(G)$ will denote the maximum degree of G.

$L(G)$ will denote the number of vertices in the largest component of G.

A *clique* in G is a complete subgraph of G, ie. a set of vertices $C \subseteq V$ such that $vw \in E$ for all $v, w \in C$.
The *clique number* $\omega(G)$ is the size of the largest clique.

A k-colouring of G is a map $f : V \to \{1, \ldots, k\}$ that satisfies $f(v) \neq f(w)$ whenever $vw \in E$
The *chromatic number* $\chi(G)$ is the least k such that G is k-colourable.

The *degeneracy* $\delta^*(G)$ is the maximum over all subgraphs of G of the minimum degree. The significance of δ^* is that if the greedy algorithm is used to colour G then $\delta^*(G) + 1$ colours will be used.

Probability terminology

If $Z_1, Z_2, \ldots$ is a sequence of random variables and c is a constant then we say that $Z_n \to c$ *in probability* if $\mathbb{P}(|Z_n - c| > \epsilon) \to 0$ for all $\epsilon > 0$.

If $A_1, A_2, \ldots$ is a sequence of events then we say that A_n holds whp. (with high probability) if $\mathbb{P}(A_n) \to 1$.

Some background: connectivity

Theorem.[Erdös-Rényi 1959]

1. If $np \leq (1-\epsilon)\ln(n)$ then $\mathbb{P}(G_n \text{ is connected }) \to 0$;
2. If $np \geq (1+\epsilon)\ln(n)$ then $\mathbb{P}(G_n \text{ is connected }) \to 1$.

Theorem.[Penrose 1998] There exists a constant c such that

1. If $nr^d \leq (c-\epsilon)\ln(n)$ then $\mathbb{P}(G_n \text{ is connected }) \to 0$;
2. If $nr^d \geq (c+\epsilon)\ln(n)$ then $\mathbb{P}(G_n \text{ is connected }) \to 1$.

Note: these are simplifications/corollaries of the actual results.

The constant c is chosen in such a way that the average degree is $\sim \ln(n)$. Some restrictions on the probability measure ν and norm $\|\,.\,\|$ apply in this slide and the one after the next. In particular ν is taken uniform on $[0,1]^d$.

nr^d and the average degree

We prefer to describe the various cases in terms of the quantity nr^d, because nr^d can be considered a measure of the average degree of the graph. Intuitively it should be obvious that the *expected* degree scales with nr^d. More formally it can be shown that:

Proposition. The average degree $\overline{d}$ of G_n satisfies:

1. If $nr^d \gg n^{-1}$ then $\frac{\overline{d}}{nr^d} \to \theta \int_{\mathbb{R}^d} f^2(x)dx$ in probability;
2. If $nr^d = O(n^{-1})$ then $\liminf_{n\to\infty} \mathbb{P}(\overline{d} = 0) > 0$.

Here θ is the d-dimensional volume of the unit ball and f is the probability density function of ν.

Some more background: the giant component

Theorem[Erdös-Rényi 1960]

1. If $np \leq 1 - \epsilon$ then $\frac{L(G(n,p))}{n} \to 0$ in probability;
2. If $np \geq 1 + \epsilon$ then $\liminf \frac{L(G(n,p))}{n} > 0$ with probability 1.

Theorem[Penrose 2003] There is a constant c such that

1. If $nr^d \leq c - \epsilon$ then $\frac{L(G_n)}{n} \to 0$ in probability;
2. If $nr^d \geq c + \epsilon$ then $\liminf \frac{L(G_n)}{n} > 0$ with probability 1.

Again these are simplifications/corollaries of the actual results.
The precise value of the constant c is unknown.

Two-point concentration in $G(n, p(n))$

Theorem.[Matula 1972] If p is fixed then

$$\mathbb{P}(\omega(G(n,p)) \in \{k(n), k(n)+1\}) \to 1,$$

where $k(n) := 2\log_b(n) - 2\log_b \log_b(n) + 1 + 2\log_b(e/2)$ with $b = 1/p$.

Theorem.[Łuczak 1991] If $p(n) \leq n^{-\frac{5}{6}-\delta}$ for some $\delta > 0$ then

$$\mathbb{P}(\chi(G(n,p)) \in \{k(n), k(n)+1\}) \to 1,$$

for some sequence $k(n)$.

Theorem.[Alon& Krivelevich 1997] The result of Łuczak also holds when $p(n) \leq n^{-\frac{1}{2}-\delta}$.

Two-point concentration in random geometric graphs: previous work

Theorem.[Månsson 1999] If M_n denotes the largest number of points contained in a ball of radius $r = r(n)$ and $nr^d = n^{-\frac{1}{k}}$ then

$$\mathbb{P}(M_n \in \{k, k+1\}) \to 1.$$

Theorem.[Penrose 2002] If $nr^d = O(1)$ then

$$\mathbb{P}(\omega(G_n) \in \{k(n), k(n)+1\}) \to 1,$$

for some sequence $k(n)$.

The result actually is shown to hold for a class of related random variables.

Two-point concentration in random geometric graphs: previous work cont'd

Theorem.[Penrose 2003] If $nr^d = o(\ln(n))$ and ν is uniform on $[0,1]^d$ then

$$\mathbb{P}(\Delta(G_n) \in \{k(n), k(n)+1\}) \to 1,$$

for some sequence $k(n)$.

He conjectured the last result to be true also for $\omega(G_n)$.

A framework: clustering rules

We assume we are given a a sequence of maps $(h_n)_n$ that assign non-negative integers to finite subsets of $\mathbb{R}^d$ s.t. for some $R_1, R_2 > 0$ the following hold:

(C1) $h_n(A) \leq |A|$;

(C2) If $A \subseteq B(x; R_1 r(n))$ for some $x \in \mathbb{R}^d$ then $h_n(A) = |A|$;

(C3) If $h_n(A) > 0$ then $\| a - b \| < R_2 r(n)$ for all $a, b \in A$;

(C4) If $h_n(A) = l$ then $h_n(A \setminus \{a\}) \geq l - 1$ for all but at most one $a \in A$.

We will be interested in the maximum $M = M(n)$ of $h_n(A)$ over all subsets $A \subseteq \{X_1, \ldots, X_n\}$. If $h_n(A) = l$ then we will say that A is an *l-cluster* and we will say that A is a *$(\geq l)$-cluster* if $h_n(A) \geq l$.

Examples

Example. We get $M(n) = \omega(G_n)$ if we set $h_n(A) = |A|$ if $diam(A) < r(n)$ and $h_n(A) = 0$ otherwise. We may take $R_1 = \frac{1}{2}, R_2 = 1$;

Example. If W is a bounded set with non-empty interior, then we can get $M(n) = \max_x |\{X_1, \ldots, X_n\} \cap (x + rW)|$, the maximum number of points contained in any translate of rW. We can put R_1 equal to the inradius and R_2 equal to the diameter of W;

Example. Set $h_n(A) = 0$ if A is not contained in some ball $B(x; r)$ and otherwise let $h_n(A)$ be equal to the maximum degree + 1 of the subgraph of G_n induced by A. Then $M(n) = \Delta(G_n) + 1$ and we can put $R_1 = \frac{1}{2}$ and $R_2 = 2$.

Today's main results

Theorem 1. If $nr^d = o(\ln n)$ and (C1)-(C4) hold then

$$\mathbb{P}(M(n) \in \{k(n), k(n)+1\}) \to 1,$$

for some sequence $k(n)$.

Corollary 2. If $r(n)$ satisfies $nr^d = o(\ln n)$ then there exists a sequence $l(n)$ such that

$$\mathbb{P}(\chi(G_n) \in \{l(n), l(n)+1\}) \to 1.$$

Corollary 3. If $r(n)$ satisfies $nr^d = o(\ln n)$ then there exists a sequence $m(n)$ such that

$$\mathbb{P}(\delta^*(G_n) \in \{m(n), m(n)+1\}) \to 1.$$

Thm. 1 proof plan

1. Guess k;
2. Let N_k be the number of points in some $(\geq k)$-cluster and let P_{k+2} be the number of pairs of points in some $(\geq k+2)$-cluster;
3. Compute $\mathbb{E}N_k, Var(N_k), \mathbb{E}P_{k+2}$ and observe:
   a) $\mathbb{P}(M < k) = \mathbb{P}(N_k = 0) \leq \mathbb{P}(|N_k - \mathbb{E}N_k| \geq \mathbb{E}N_k) \leq Var(N_k)/(\mathbb{E}N_k)^2 = o(1)$;
   b) $\mathbb{P}(M > k+1) = \mathbb{P}(P_{k+2} \geq \binom{k+2}{2}) = O(\mathbb{E}P_{k+2}/k^2) = o(1)$.

Tools

The only fact about ν that we need is that f is bounded.

Lemma. There exists a constant $\eta > 0$, dependent only on ν, such that the following holds. For any bounded set $W \subseteq \mathbb{R}^d$ with $\lambda(W) > 0$ and any $r > 0$ there exist $\Omega(r^{-d})$-many disjoint translates $W_1, \ldots, W_K$ of rW with $\nu(W_i)/\lambda(W_i) > \eta$ for all i.

Tools

Lemma. Let $Z \sim Bi(n, p)$ and $k \geq \mu := np$. Then

$$(\frac{\mu}{ek})^k \leq \mathbb{P}(Z \geq k) \leq (\frac{e\mu}{k})^k$$

Lemma. Let $(Z_1, \ldots, Z_m) \sim mult(n; p_1, \ldots, p_m)$. Then

$$\mathbb{P}(Z_1 \leq k_1, \ldots, Z_m \leq k_m) \leq \Pi_{i=1}^m \mathbb{P}(Z_i \leq k_i).$$

Slightly more detailed sketch

Set $j = j(n) = \ln n / \ln(\frac{\ln n}{nr^d})$. (Note $j/nr^d \to \infty$).

Lemma. $(1-\epsilon)j \leq M \leq (1+\epsilon)j + 1$ whp.

Proof sketch:

1. $\mathbb{P}(M \leq (1-\epsilon)j) \leq \mathbb{P}(Bi(n, const \cdot r^d) \leq (1-\epsilon))^{\Omega(r^{-d})} \leq (1 - (\frac{const \cdot nr^d}{j})^{(1-\epsilon)j})^{\Omega(r^{-d})}$
2. $\mathbb{P}(M \geq (1+\epsilon)j + 1) \leq n \cdot \mathbb{P}(Bi(n, const \cdot r^d) \geq (1+\epsilon)j) \leq n(\frac{const \cdot nr^d}{j})^{(1+\epsilon)j}$.

Note if $nr^d \leq n^{-c}$ then j remains bounded and we're done. So we assume $nr^d \geq n^{-\frac{3}{2}}$ in the sequel.

Slightly more detailed sketch cont'd

Now choose k such that

$$\mathbb{E}N_k - k(\frac{j}{nr^d})^{\frac{1}{4}} \geq 0 > \mathbb{E}N_{k+1} - (k+1)(\frac{j}{nr^d})^{\frac{1}{4}},$$

where N_k is the number of points in clusters of size at least k.

Lemma. For n suff. large $k \geq \frac{1}{2}j$.

Lemma. $Var(N_k) = O(k\mathbb{E}N_k)$.

This lemma contains the bulk of the computations.

We can conclude

$$Var(N_k)/(\mathbb{E}N_k)^2 = O(k/\mathbb{E}N_k) = O((\frac{j}{nr^d})^{-\frac{1}{4}}) = o(1).$$

Slightly more detailed sketch cont'd

Now consider P_{k+2}.

Lemma. $\mathbb{E}P_{k+2} = O(nr^d \mathbb{E}N_{k+1})$.

Proof sketch: The probability that X_1, X_2 are in a common $(\geq k+2)$-cluster is at most twice the probability that X_1 is in some $(\geq k+1)$-cluster and $\|X_1 - X_2\| \leq R_2 r$, using (C4).

So, $\mathbb{E}M_{k+2}/k^2 = O(k(\frac{j}{nr^d})^{\frac{1}{4}} \cdot nr^d/k^2) = O((\frac{nr^d}{k})(\frac{j}{nr^d})^{\frac{1}{4}}) = O((\frac{j}{nr^d})^{-\frac{3}{4}} = o(1)$.

This concludes the proof (sketch) of thm. 1.

Cor. 1 proof sketch

As part of the proof of thm. 1 we have seen that for any $\epsilon > 0$:

$$(1-\epsilon)j \leq M \leq (1+\epsilon)j + 1 \text{ whp.},$$

where $j = \ln n / \ln(\frac{\ln n}{nr^d})$. So in particular $(1-\epsilon)j \leq k \leq (1+\epsilon)j$, regardless of the clustering rule that was chosen.

Fix K (large),ϵ (small). Set $M :=$ max. chromatic number of a subgraph induced by the points in a ball of radius Kr. Whp. the following hold:

(i) $M \in \{k, k+1\}$;

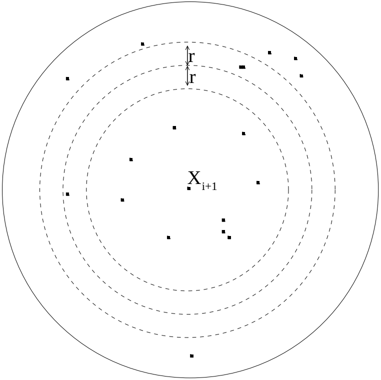
(ii) No ball of radius Kr contains more than $(1+\epsilon)j + 1$ points.

Cor. 1 proof sketch cont'd

It suffices to show that if (i),(ii) hold then the chromatic number is M. Set $V_i := \{X_1, \ldots, X_i\}$ and suppose we can colour the subgraph $G_n[V_i]$ induced by V_i with M colours. We can also colour $G_n[V_{i+1} \cap B(X_{i+1}; Kr)]$ with M colours.

As K is large, ϵ small the pigeon hole will give us some "annulus" $A = B(X_{i+1}; rm) \setminus B(X_{i+1}; r(m-2))$ with no more than M points.

Cor. 1 proof sketch cont'd



There exist an M-colouring of $V_{i+1} \setminus B(X_{i+1}; r(m-1))$ and of $V_{i+1} \cap B(X_{i+1}; r(m-1))$. Permute these colourings in such a way that they use disjoint sets of colours on the annulus A.

What happens for other r?

Theorem. If ν is uniform on $[0,1]^d$ then the following hold:

(i) If $nr^d = o(\ln(n))$ then there is a sequence $a(n)$ such that

$$\mathbb{P}(\Delta(G_n) \in \{a(n), a(n)+1\}) \to 1;$$

(ii) If $nr^d = t\ln(n)$ then there is a constant $\beta = \beta(t) > 1$ and sequences $a(n), \gamma(n)$ with $\beta^{-1} \leq \gamma(n) \leq 1$ s.t.:

$$\mathbb{P}(\Delta(G_n) < a(n) + K) \sim e^{-\gamma(n)\beta^{-K}},$$

for any fixed K;

(iii) If $\ln(n) \ll nr^d \ll \ln(n)^d$ then there exist sequences $a(n), b(n)$ such that:

$$\mathbb{P}\left(\frac{\Delta(G_n) - a(n)}{b(n)} < x\right) \to e^{-e^{-x}},$$

for all $x \in \mathbb{R}$.

Part one is already due to Penrose.

Proof sketch

1. Let N_k denote the number of points of degree at least k;
2. For k in the appropriate range, N_k is approximately Poisson (Stein's method);
3. So, $\mathbb{P}(\Delta < k) = \mathbb{P}(N_k = 0) \approx e^{-\mathbb{E}N_k}$;
4. It now suffices to compute the expectations $\mathbb{E}N_k$.

Further work

1. Other ranges of r;
2. The exact value of k in the range considered;
3. Maybe less ambitiously: do the differences $\chi(G_n) - \omega(G_n)$ stay bounded?