Isolated Vertices of Random Niche Overlap Graphs

Zbigniew J. Palka,[1] Joel E. Cohen

Rockefeller Univerity
1230 York Avenue
New York, NY 10021-6399

Institute for Advanced Study
Olden Lane
Princeton, NJ 08540, U.S.A.

1. Introduction.

1.1 Definitions

The vertex set of random graphs and random digraphs will be denoted $[n] = \{1, 2, \ldots, n\}$. For any distinct $i, j \in [n]$, $\{i, j\}$ will denote the (undirected) edge between vertices i and j of a graph on $[n]$. Loops, that is, edges of the form $\{i, i\}$, are excluded. For $0 \leq p \leq 1$, G_p will denote the random graph on $[n]$ such that each possible edge occurs independently with identical probability p.

For any distinct $i, j \in [n]$, (i, j) will denote the arc from vertex i to vertex j of a digraph on $[n]$. The vertex i will be said to be the tail of the arc and the vertex j will be said to be the head of the arc. Loops, that is, arcs of the form (i, i), are excluded. For $0 \leq p \leq 1$, W_p will denote the random digraph on $[n]$ obtained from G_p by orienting each edge $\{i, j\}$ of G_p from the vertex with the smaller number to the vertex with the larger number; that is, $\{i, j\}$ is an edge of G_p if and only if $(\min(i, j), \max(i, j))$ is an arc of W_p. In the special case where $p = c/n$ and c is a positive constant independent of n, W_p has been called the cascade model (Cohen and Newman 1985). The niche overlap graph (often called the competition graph) $G(W)$ of any digraph W without loops is a graph on $[n]$ that has an edge $\{i, j\}$ between distinct vertices $i, j \in [n]$ if and only if there is a third distinct vertex $k \in [n]$ such that (k, i) and (k, j) are arcs of W. The random niche overlap graph $G(W_p)$ of W_p differs considerably from the ordinary random graph model because, in $G(W_p)$, the probability of an edge between two vertices depends on their labels:

$$P\left(\{i, j\} \text{ is an edge of } G(W_p)\right) = 1 - (1 - p^2)^{i-1},$$

[1] Permanent address: Department of Discrete Mathematics, Adam Mickiewicz University, Matejki 48/49, 60-769 Poznań, Poland.

$i = 1, 2, \ldots, n-1$ and $i < j$, and more importantly, these probabilities are strongly correlated. This makes the investigation of $G(W_p)$ more complex than of G_p. Henceforth, we call $G(W_p)$ a random overlap graph.

In this note, some preliminary results about the number X_0 of vertices of degree zero in $G(W_p)$ are presented. The mean and variance of X_0 over the entire range of behaviors of p as a function on n, both for finite n and in the limit as n becomes large, are computed. The interesting question about the asymptotic probability distribution of X_0 remains open.

Investigations of a random overlap graph model have started very recently. Thus, it seems appropriate to present results obtained up to now. We do it in the next paragraph.

(Though it is not necessary from the mathematical point of view, it may be helpful to the intuition to sketch very briefly the ecological interpretation of some of these concepts. In models of food webs in community ecology, the vertices correspond to species or other kinds of organisms. An arc (i, j) means that j eats i, since material and energy flow from i to j. There is an edge $\{i, j\}$ in the niche overlap graph if and only if the diets of species i and j overlap, that is, if and only if there is potential competition between i and j [hence, the earlier term, competition graph]. The niche overlap graph is a combinatorial caricature of the structure of dietary overlaps or potential competition in an ecosystem.)

1.2 Some previous results

The random overlap graph $G(W_p)$ has interesting properties that differ notably from those of the usual random graph G_p.

The threshold function of $G(W_p)$ for the appearance of a complete subgraph on $m \geq 3$ vertices is $n^{-1-1/m}$ (Cohen and Palka 1990) whereas the corresponding function for G_p is $n^{-2/(m-1)}$ (Erdös and Rényi 1960). Thus, the first triangle of G_p appears (with a positive probability) when $p = c/n$ where $0 < c < \infty$. However, the corresponding overlap graph $G(W_p)$ already contains a complete subgraph on approximately $\log n / \log\log n$ vertices (Luczak and Palka manuscript).

Now consider the property of being interval. A graph on $[n]$ is interval when there is a collection $I_1, I_2, \ldots, I_n$ of intervals of the real line such that there is an edge between i and j $(i \neq j)$ if and only if I_i and I_j overlap. Assume that $p = p(n) \to 0$ so that $pn^{7/6} = c$. Then (Cohen, Komlós, Mueller 1979) almost every G_p is interval if $c = c(n) \to 0$ and is not interval if $c = c(n) \to \infty$. On the threshold, that is, when $0 < c < \infty$,

$$\lim_{n \to \infty} P(G_p \text{ is interval}) = e^{-\lambda}.$$

where $\lambda = c^6/6$. In the case of a random overlap graph the following result holds (Cohen and Palka 1990). Let $p = p(n) \to 0$ so that $pn^{10/9} = d$. Then almost every $G(W_p)$ is interval if $d = d(n) \to 0$ and is not interval if $d = d(n) \to \infty$.

When $0 < d < \infty$, then

$$\lim_{n\to\infty} P\left(G(W_p) \text{ is interval}\right) = e^{-\mu},$$

where $\mu = (9170/10!)\,d^9$.

Finally, the triangularity of $G(W_p)$ with p around c/n behaves as follows (Cohen and Palka 1990). (A graph is triangulated if it has no induced cycles of four or more edges.) Let $p = c/n$, $0 < c < \infty$. Then

$$\lim_{n\to\infty} P\left(G(W_p) \text{ is triangulated}\right) = e^{-\gamma},$$

where

$$\gamma = \sum_{\substack{k=8 \\ k \text{ even}}} \frac{c^k a(k)}{k!}$$

and

$$a(k) = (-4)^{k/2} \sum_{m=1}^{k-1} (-1)^m m!\, S(k-1, m)\, 2^{-2-m}$$

where $S(k, m)$ are Stirling's numbers of the second kind.

2. Vertex degrees.

2.1 Expected number of isolated vertices in the overlap graph

We distinguish two types of vertices of W_p that are isolated vertices of $G(W_p)$. We say that vertex i is of type 1 if *in-deg*$(i) = 0$. Vertex i is said to be of type 2 if *in-deg*$(i) \geq 1$ and for every k such that (k, i) is an arc, (k, j) is not an arc for $j = k+1, \ldots, n$ and $j \neq i$. Then vertex i of $G(W_p)$ is isolated if and only if the corresponding vertex i of W_p is either of type 1 or of type 2. Thus, for $1 \leq i \leq n$, if $q = 1 - p$, then

$$P(i \text{ is isolated}) = \prod_{k=1}^{i-1} (q + pq^{n-k-1}) = q^{i-1} a_{i-1}$$

where

$$a_{i-1} = \prod_{k=1}^{i-1} (1 + pq^{n-k-2}), \quad 2 \leq i \leq n. \qquad (2.1.1)$$

Consequently, the expected number of isolated vertices is

$$E(X_0) = \sum_{i=1}^{n} q^{i-1} a_{i-1}. \qquad (2.1.2)$$

Assume that $p \le 1/2$. Then $pq^{n-k-2} \le 1$ for $1 \le k \le n-1$ and by the fact $1+x = \exp\{x + O(x^2)\}$, $|x| \le 1$, we obtain

$$a_{i-1} = \exp\left\{\sum_{k=1}^{i-1} [pq^{n-k-2} + O(p^2 q^{2(n-k-2)})]\right\} = e^{h(i-1)},$$

where

$$h(i-1) = q^{n-i-1} - q^{n-2} + O\left[\frac{p}{1+q}(q^{2(n-i-1)} - q^{2(n-2)})\right] \qquad (2.1.3)$$

for $1 \le i \le n$. When $p > 1/2$, we cannot apply the above approach for $i = n$. Thus, we separate the last term of the summation in (2.1.2). This quantity equals

$$q^{n-1} a_{n-1} = q^{n-1}(1 + p/q)e^{h(n-2)} = O(q^{n-2}) = O(e^{-n/2}) \qquad (2.1.4)$$

and is insignificant (see Property 2.1.8 (b)).

Now we examine the behavior of the expected number of isolated vertices of $G(W_p)$ when p changes from zero to one. We consider three intervals for the values of p.

Property 2.1.5. *Let* $p = (n\omega(n))^{-1}$ *where* $\omega(n) \to \infty$ *as* $n \to \infty$. *Then*

$$E(X_0) = \frac{1-q^n}{p}[1 + O(1/\omega(n))].$$

Proof: By (2.1.3), $h(i-1) = O(1/\omega(n))$. Thus, using the fact $e^x = 1 + O(x)$, $|x| \le 1$, we obtain our result directly from (2.1.2). ∎

Property 2.1.6. *Let* $np = c$ *where* $0 < c < \infty$. *Then for sufficiently large* n

$$E(X_0) = \frac{n}{c}e^{-c}e^{-e^{-c}}[e^c - 1 + c + f(c)][1 + O(1/n)],$$

where

$$(1 - e^{-c})/2 + (1 - e^{-2c})/12 \le f(c) \le (e-2)(1 - e^{-c}).$$

Proof: By (2.1.3), $h(i-1) = q^{n-i-1} - q^{n-2} + O(1/n)$ for $1 \le i \le n$. Thus, by (2.1.2),

$$E(X_0) = \exp\{-q^{n-2} + O(1/n)\}W(n,p)$$

where

$$W(n,p) = \sum_{i=1}^{n} q^{i-1}\exp\{q^{n-i-1}\}.$$

Clearly,

$$
\begin{aligned}
W(n,p) &\geq \sum_{i=1}^{n} q^{i-1}\left(1+q^{n-i-1}+\frac{1}{2}q^{2(n-i-1)}+\frac{1}{6}q^{3(n-i-1)}\right) \\
&= \frac{1-q^n}{p}+nq^{n-2}+\frac{1}{2}q^{n-3}\frac{1-q^n}{1-q}+\frac{1}{6}q^{n-4}\frac{1-q^{2n}}{1-q^2}.
\end{aligned}
\tag{2.1.7}
$$

On the other hand,

$$
\begin{aligned}
W(n,p) &= \sum_{k=0}^{\infty}\frac{1}{k!}\sum_{i=1}^{n} q^{k(n-i-1)+(i-1)} \\
&= \frac{1-q^n}{p}+nq^{n-2}+\frac{q^n}{p}\sum_{k=2}^{\infty}\frac{1}{k!}R_k
\end{aligned}
$$

where

$$
R_k = R_k(n,p) = \left(\frac{1}{q}\right)^{k+1}\frac{1-q^{(k-1)n}}{1+q+\ldots+q^{k-2}}.
$$

If $p = c/n$, where c is a constant, then $R_k \leq R_2$ for all $3 \leq k \leq n$. Thus, for sufficiently large n,

$$
W(n,p) \leq \frac{1-q^n}{p}+nq^{n-2}+\frac{q^n}{p}\frac{1-q^n}{q^3}(e-2).
$$

This, together with (2.1.7), implies our result. ∎

One can improve slightly the lower bound of $E(X_0)$ by including more terms in (2.1.7). When c is close to zero, then $f(c)$ is also close to zero. However, the gap between the lower and upper bound for $E(X_0)$ increases with c. In the worst case (when c is sufficiently large), we have $1/2 \leq f(c) \leq e-2$. Thus, when $c = c(n) \to \infty$ as $n \to \infty$, we use a different approach which gives a precise result.

Property 2.1.8. *(a) Let $np = c(n)$ where $c(n) \to \infty$ as $n \to \infty$ and $c(n) = o(n)$. Then*

$$
E(X_0) = \frac{1}{p}\left[1+O(e^{-\alpha(n)})\right]
$$

where $\alpha(n) = c(n)/\gamma(n)$ and $\gamma(n)$ is an arbitrary sequence tending to infinity such that $\gamma(n) = o(c(n))$.

(b) *Let* $0 < p \leq 1$ *be a constant. Then*

$$\lim_{n \to \infty} E(X_0) = \frac{1}{p}.$$

Proof: (a) Let $m = n - n/\gamma(n)$. Then, by (2.1.3), for $1 \leq i \leq m$

$$h(i-1) = O(e^{-\alpha(n)})$$

and $0 \leq h(i-1) \leq 1 + o(1)$ for $m+1 \leq i \leq n$. Thus, with $\beta(n) = 1 + O(e^{-\alpha(n)})$,

$$\sum_{i=1}^{m} q^{i-1} e^{h(i-1)} = \frac{1-q^m}{p}\beta(n) = \frac{1}{p}\beta(n) \tag{2.1.9}$$

since $q^m = O[\exp\{\alpha(n) - c(n)\}]$. Furthermore,

$$\sum_{i=m+1}^{n} q^{i-1} e^{h(i-1)} = O(q^m/p), \tag{2.1.10}$$

which completes the proof of (a).

(b) If $0 < p \leq 1/2$, then the proof follows the same lines as in part (a). When $p > 1/2$ then, by (2.1.4), the last term in (2.1.10) is $O(e^{-n/2})$ and the result follows as before. ∎

2.2 Variance of the number of isolated vertices

We now estimate the second factorial moment $E_2(X_0)$ of the number of isolated vertices X_0. The probability p_{ij} that two given vertices, say i and j where $i < j$, are isolated in $G(W_p)$ equals

$$\begin{aligned} p_{ij} &= \prod_{k=1}^{i-1}(q^2 + 2pq^{n-k-1}) \prod_{k=i}^{j-1}(q + pq^{n-k-1}) \\ &= q^{i-1} b_{i-1} q^{j-1} a_{j-1} \end{aligned}$$

where

$$b_{i-1} = \prod_{k=1}^{i-1} \frac{1 + 2pq^{n-k-3}}{1 + pq^{n-k-2}}$$

and a_{j-1} is given by (2.1.1). Again, by using $1 + x = \exp\{x + O(x^2)\}$, a routine calculation shows that for $p \leq 1/3$ and $2 \leq i \leq n-1$,

$$b_{i-1} = \exp\left\{(2-q)(1-q^{i-1})q^{n-i-2} + O\left[\frac{p}{1+q}(1 - q^{2(i-1)})q^{2(n-i-2)}\right]\right\}. \tag{2.2.1}$$

When $p > 1/3$ and $i = n-1$,

$$b_{n-2} = b_{n-3}\,\frac{1+2p/q}{1+p} = \exp\{O(1)\}.$$

Now, by (2.1.2),

$$\begin{aligned} E_2(X_0) &= 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p_{ij} \\ &= 2\sum_{i=1}^{n-1} q^{i-1} b_{i-1}\left(E(X_0) - \sum_{j=1}^{i} q^{j-1} a_{j-1}\right) \end{aligned} \tag{2.2.2}$$

and we are ready to estimate $E_2(X_0)$ for the entire range of p. This time we begin with the case when $np = c(n)$ where $c(n) \to \infty$ as $n \to \infty$.

Property 2.2.3. *Under the assumptions of Property 2.1.8 we have*

$$E_2(X_0) = \frac{2q}{p^2(1+q)}[1 + O(e^{-\alpha(n)})].$$

Proof: We use the same notation as in the proof of Property 2.1.8. Thus, for $1 \leq i \leq m$, we have (compare (2.1.9))

$$\sum_{j=1}^{i} q^{j-1} a_{j-1} = \frac{1-q^i}{p}\beta(n).$$

Also, by (2.2.1),

$$b_{i-1} = \exp\{O(e^{-\alpha(n)})\} = \beta(n).$$

Consequently, by (2.2.2),

$$E_2(X_0) = 2\beta(n)\sum_{i=1}^{m} q^{i-1}\left(E(X_0) - \frac{1-q^i}{p}\beta(n)\right) + S_m$$

where

$$S_m = 2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^{n} p_{ij}.$$

Applying Property 2.1.8 we get

$$\begin{aligned} E_2(X_0) &= \frac{2}{p}\beta(n)\sum_{i=1}^{m} q^{i-1}\left(q^i + O(e^{-\alpha(n)})\right) + S_m \\ &= \frac{2q}{p^2(1+q)}\beta(n) + S_m. \end{aligned} \tag{2.2.4}$$

Furthermore, for $2 \leq i \leq n-1$, $b_{i-1} = \exp\{O(1)\}$. Consequently, by (2.1.10),

$$S_m = O\left(\frac{q^m}{p}\sum_{i=m+1}^{n-1} q^{i-1}\right) = O\left(\frac{1}{p^2}e^{2[\alpha(n)-c(n)]}\right).$$

This, together with (2.2.4), implies the property. ∎

Corollary 2.2.5. *Under the assumptions of Property 2.1.8,*

$$Var(X_0) = \frac{q}{p(1+q)}[1 + O(e^{-\alpha(n)})].$$

When $np = c$, where c is a positive constant, we present only the following crude bound for $E_2(X_0)$.

Property 2.2.6. *Let $p = c/n$ where $0 < c < \infty$. Then*

$$E_2(X_0) \leq \frac{2n}{c}E(X_0)\exp\{1 - e^{-c} + o(1)\}(1 - e^{-c}).$$

Proof: By (2.2.1), for $1 \leq i \leq n-1$,

$$b_{i-1} \leq \exp\{(2-q)(1-q^{n-1})q^{-1} + O(1/n)\}.$$

Thus, by (2.2.2)

$$E_2(X_0) \leq 2\sum_{i=1}^{n-1} q^{i-1} b_{i-1} E(X_0).$$

Since $q = 1 - O(1/n)$, taking n sufficiently large yields the desired bound. ∎

Finally, when $np = o(1)$ we have

Property 2.2.7. *Let $p = (n\omega(n))^{-1}$ where $\omega(n) \to \infty$. Then*

$$E_2(X_0) = E(X_0)^2(1 + o(1/n)).$$

Proof: The equality follows directly from (2.2.2) since both b_{i-1} and a_{j-1} are of the order $1 + O(1/\omega(n))$. ∎

Corollary 2.2.8. *Under the assumption of Property 2.2.7,*

$$Var(X_0) = E(X_0) + o\left(\frac{1}{n}E(X_0)^2\right).$$

2.3 *Distribution of the number of isolated vertices*

The interesting question about the asymptotic probability distribution of the random variable X_0 remains open. We remark only that in a case when p does not depend on n,

$$P(X_0 \geq 2) \geq c(p)$$

where $c(p)$ is a positive constant depending on p. For example, when $p = \frac{1}{2}$ then, for large n,

$$P(X_0 \geq 2) \geq \frac{3}{5}. \tag{2.3.1}$$

As a matter of fact, let Y_0 be the number of isolates among vertices $\{2, 3, \ldots, n\}$ in $G(W_p)$. From the strong second moment inequality

$$P(Y_0 \geq 1) \geq \frac{E^2(Y_0)}{E(Y_0^2)}.$$

Since $Y_0 = X_0 - 1$ so taking into account Property 2.1.8 (b) and Corollary 2.2.5 one gets the bound (2.3.1).

2.4 *Vertices with higher degrees*

The vertices of degree $k \geq 1$ in $G(W_p)$ are more difficult to study. Even when k is small, it is hard to keep track of all configurations in W_p that contribute to vertices of degree k in $G(W_p)$. We present here only a formula for the expected number of vertices of degree one in $G(W_p)$.

For vertex i, what is the probability $p_i(j)$ that its only neighbor in $G(W_p)$ is a given vertex j? Let $t = \min\{i, j\}$ and $h = t - 1$. Consider vertices of W_p below i and j. Then a vertex k $(1 \leq k \leq h)$ is one of three types: "good", if k is connected to i and to j only; "bad", if k is connected to i and to j, plus at least one other vertex; "harmless", if k is not connected to both i and j. Clearly,

$$P(k \text{ is "good"}) = p^2 q^{n-k-2},$$
$$P(k \text{ is "bad"}) = p^2(1 - q^{n-k-2}),$$

and

$$P(k \text{ is "harmless"}) = 1 - p^2.$$

If vertex i has degree 1, at least one of the h vertices below i and j must be "good" and all other vertices must be "harmless". In other words, each of h vertices has to be "good" or "harmless" but not all of them can be "harmless". Thus,

$$p_i(j) = \prod_{k=1}^{h}(p^2 q^{n-k-2} + 1 - p^2) - (1 - p^2)^h.$$

Consequently, the expected number of vertices X_1 of degree 1 is

$$E(X_1) = \sum_{i=2}^{n-1} \sum_{\substack{j=2 \\ j \neq i}}^{n} p_i(j).$$

In general, formulas for $E(X_k)$ become awkward when k gets large.

Acknowledgement.

This work was supported in part by U.S. National Science Foundation grant BSR 87-05047, and by the hospitality of Mr. and Mrs. William Golden.

References

1. B. Bollobás, "Random Graphs", Academic Press, New York, 1985.
2. J.E. Cohen, "Food Webs and Niche Space", 189 pp, Princeton University Press, Princeton, 1978.
3. J.E. Cohen, and C.M. Newman, *A stochastic theory of community food webs: I. Models and aggregated data*, Proceedings of the Royal Society (London) Series B **224** (1985), 421–448.
4. J.E. Cohen, J. Komlós, and T. Mueller, *The probability of an interval graph and why it matters*, Proc. Symp. Pure Math. **34** (1979), 97–115.
5. J.E. Cohen and Z.J. Palka, *A stochastic theory of community food webs: V. Intervality and triangulation in the trophic niche overlap graph*, American Naturalist **135** (1990), 435–463.
6. P. Erdös and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5** (1960), 17–61.
7. T. Luczak and Z.J. Palka, *On the clique number of a random overlap graph*, Manuscript.
8. F. Roberts (ed.), "Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. IMA Vol. 17", Springer-Verlag, New York, 1989.