A note on network reliability

Noga Alon *

Institute for Advanced Study, Princeton, NJ 08540

and Department of Mathematics

Tel Aviv University, Tel Aviv, Israel

Let $G = (V, E)$ be a loopless undirected multigraph, with a probability p_e, $0 \leq p_e \leq 1$ assigned to every edge $e \in E$. Let G_p be the random subgraph of G obtained by deleting each edge e of G, randomly and independently, with probability $q_e = 1 - p_e$. For any nontrivial subset $S \subset V$ let $(S, \overline{S})$ denote, as usual, the cut determined by S, i.e., the set of all edges of G with an end in S and an end in its complement $\overline{S}$. Define $P(S) = \sum_{e \in (S, \overline{S})} p_e$, and observe that $P(S)$ is simply the expected number of edges of G_p that lie in the cut $(S, \overline{S})$. In this note we prove the following.

Theorem 1 *For every positive constant b there exists a constant $c = c(b) > 0$ so that if $P(S) \geq c \log n$ for every nontrivial $S \subset V$, then the probability that G_p is disconnected is at most $1/n^{-b}$.*

The assertion of this theorem (in an equivalent form) was conjectured by Dimitris Bertsimas, who was motivated by the study of a class of approximation graph algorithms based on a randomized rounding technique of solutions of appropriately formulated linear programming relaxations. Observe that the theorem is sharp, up to the multiplicative factor c, by the well known results on the connectivity of the random graph (see, e.g., [2]). In case our G

*This research was supported in part by the Institute for Mathematics and its Applications with funds provided by the NSF and by the Sloan Foundation, Grant No. 93-6-6. The paper appeared in: Discrete Probability and Algorithms (D. Aldous, P. Diaconis, J. Spencer and J. M. Steele eds.), IMA Volumes in Mathematics and its applications, Vol. 72, Springer Verlag (1995), 11-14.

above is simply the complete graph on n vertices, and $p_e = p$ for every edge e, these known results assert that the subgraph G_p, which in this case is simply the random graph $G_{n,p}$, is almost surely disconnected if $p = (1 - \epsilon) \log n/n$, although in this case $P(S) = \Omega(\log n)$ for all S. Theorem 1 can thus be viewed as a generalization to the case of non-uniform edge probabilities of the known fact that if $p > (1 + \epsilon) \log n/n$ then the random graph $G_{n,p}$ is almost surely connected. It would be interesting to extend some other similar known results in the study of random graphs to the non-uniform case and obtain analogous results for the existence of a Hamilton cycle, a perfect matching or a k-factor.

The above theorem is obviously a statement on network reliability. Suppose G represents a network that can perform iff it is connected. If the edges represent links and the failure probability of the link e is q_e, then the probability that G_p remains connected is simply the probability that the network can still perform. The network is reliable if this probability is close to 1. Thus, the theorem above supplies a sufficient condition for a network to be reliable, and this condition is nearly tight in several cases.

Proof of Theorem 1. Let $G = (V, E)$ be a loopless multigraph and suppose that $P(S) \geq c \log n$ for every nontrivial $S \subset V$. It is convenient to replace G by a graph G' obtained from G by replacing each edge e by $k = c \log n$ parallel copies with the same endpoints and by associating each copy e' of e with a probability $p'_{e'} = p_e/k$. For every nontrivial $S \subset V$, the quantity $P'(S)$ defined by $P'(S) = \sum_{e' \in (S, \overline{S})} p'_{e'}$ clearly satisfies $P'(S) = P(S)$. Moreover, for every edge e of G, the probability that no copy e' of e survives in $G'_{p'}$ is precisely $(1 - p_e/k)^k \geq 1 - p_e$ and hence G_p is more likley to be connected than $G'_{p'}$. It therefore suffices to prove that $G'_{p'}$ is connected with probability at least $1 - n^{-b}$. The reason for considering G' instead of G is that in G' the edges are naturally partitioned into k classes, each class consisting of a single copy of every edge of G. Our proof proceeds in phases, starting with the trivial spanning subgraph of G' that has no edges. In each phase we randomly pick some of the edges of G' that belong to a fresh class which has not been considerd before, with the appropriate probability. We will show that with high probability the number of connected components of the subgraph of G' constructed in this manner decreases by a constant factor in many phases until it becomes 1, thus forming a connected subgraph. We need the following simple lemma.

Lemma 2 *Let $H = (U, F)$ be an arbitrary loopless multigraph with a probability w_f assigned to each of its edges f, and suppose that for every vertex u of H, $\sum_{v \in U, uv \in E} w_{uv} \geq 1$. Let H_w be the random subgraph of H obtained by deleting every edge f of H, randomly and independently, with probability $1 - w_f$. Then, if $|U| > 1$, with probability at least $1/2$ the number of connected components of H_w is at most $(1/2 + 1/e)|U| < 0.9|U|$.*

Proof. Fix a vertex u of H. The probability that u is an isolated vertex of H_w is precisely

$$\prod_{v \in U, uv \in E} (1 - w_{uv}) \leq exp\{-\sum_{v \in U, uv \in E} w_{uv}\} \leq 1/e.$$

By linearity of expectation, the expected number of isolated vertices of H_w does not exceed $|U|/e$, and hence with probability at least $1/2$ it is at most $2|U|/e$. But in this case the number of connected components of H_w is at most

$$2|U|/e + \frac{1}{2}(|U| - 2|U|/e) = (1/2 + 1/e)|U|,$$

as needed. □

Returning to our graph G and the associated graph G', let $E_1 \cup E_2 \cdots \cup E_k$ denote the set of all edges of G', where each set E_i consists of a single copy of each edge of G. For $0 \leq i \leq k$, define G'_i as follows. G'_0 is the subgraph of G' that has no edges, and for all $i \geq 1$, G'_i is the random subgraph of G' obtained from G'_{i-1} by adding to it each edge $e' \in E_i$ randomly and independently, with probability $p'_{e'}$. Let C_i denote the number of connected components of G'_i. Note that as G'_0 has no edges $C_0 = n$ and note that G'_k is simply $G'_{p'}$. Let us call the index i, $(1 \leq i \leq k)$, *successful* if either G'_{i-1} is connected (i.e., $C_{i-1} = 1$) or if $C_i < 0.9C_{i-1}$.

Claim: For every index i, $1 \leq i \leq k$, the conditional probability that i is successful given any information on the previous random choices made in the defintion of G'_{i-1} is at least $1/2$.

Proof: If G_{i-1} is connected then i is successful and there is nothing to prove. Otherwise, let $H = (U, F)$ be the graph obtained from G'_{i-1} by adding to it all the edges in E_i and by contracting every connected component of G'_{i-1} to a single vertex. Note that since $P'(S) \geq c \log n = k$ for every nontrivial S it follows that for every connected component D

of G'_{i-1}, the sum of probabilities associated to edges $e \in E_i$ that connect vertices of D to vertices outside D is at least 1. Therefore, the graph H satisfies the assumptions of Lemma 2 and the conclusion of this lemma implies the assertion of the claim. □

Observe, now, that if $C_k > 1$ then the total number of successes is stricly less than $\log n / \log 0.9$ ($< 10 \log_e n$). However, by the above claim, the probability of this event is at most the probability that a Binomial random variable with parameters k and 0.5 will attain a value of at most $r = 10 \log_e n$. (The crucial observation here is that this is the case despite the fact that the events "i is successful" for differnet values of i are *not* independent, since the claim above places a lower bound on the probability of success given any previous history.) Therefore, by the standard estimates for Binomial distributions (c.f., e.g., [1], Appendix A, Theorem A.1), it follows that if $k = c \log n = (20 + t) \log_e n$ then the probability that $C_k > 1$ (i.e., that $G'_{p'}$ is disconnected) is at most $n^{-t^2/2c}$, completing the proof of the theorem. □

Remarks

1. The assertion of Lemma 2 can be strengthened and in fact one can show that there are two positive constants c_1 and c_2 so that under the assumptions of the lemma the number of connected components of the random subgraph H_w is at most $(1 - c_1)|U|$ with probability at least $1 - e^{-c_2 m}$. This can be done by combining the Chernoff bounds with the following simple lemma, whose proof is omitted

   Lemma 3 *Let $H = (U, F)$ be an arbitrary loopless multigraph with a non-negative weight w_e associated to each of its edges e. Then there is a partition of $U = U_1 \cup U_2$ into two disjoint subsets so that for $i = 1, 2$ and for every vertex $u \in U_i$,*

   $$\sum_{uv \in E,\ v \in U_{3-i}} w_{uv} \geq \frac{1}{2} \sum_{uv \in E,\ v \in U} w_{uv}.$$

   For our purposes here the weaker assertion of Lemma 2 suffices.

2. It is interesting to note that several natural analogs of Theorem 1 for other graph properties besides connectivity are false. For example, it is not difficult to give an example of a graph $G = (V, E)$ and a probability function p, together with two distinguished vertices s and t, so that $P(S) \geq \Omega(n/\log n)$ ($>> \Omega(\log n)$) for all cuts S separating s and t and yet in the random subgraph G_p almost surely s and t lie in different

connected components. A simple example showing this is the graph G consisting of $n/10\log n$ internally vertex disjoint paths of length $10\log n$ each between s and t, in which $p_e = 1/2$ for every edge e.

3. Another, more interesting example showing that a natural analog of Theorem 1 for bipartite matching fails is the following. Let A and B be two disjoint vertex classes of cardinality n each. Let A_1 be a subset of $c_1 n$ vertices of A and let B_1 be a subset of $c_1 n$ vertices of B, where, say, $1/8 < c_1 < 1/4$. Let H_1 be the bipartite graph on the classes of vertices A and B in which every vertex of A_1 is connected to every vertex of B and every vertex of B_1 is connected to every vertex of A. Let H_2 be a bipartite constant-degree expander on the classes of vertices A and B; for example, a C_2-regular graph so that between any two subsets X of A and Y of B containing at least $c_1 n/2$ vertices each there are at least $c_1 n$ edges (it is easy to show that such a graph exists using a probabilistic construction, or some of the known constructions of explicit expanders). Finally, let $H = (V, E)$ be the bipartite graph on the classes of vertices A and B whose edges are all edges of H_1 or H_2. Define, also, $p_e = 1/(4C_2)$ for every edge e of H. It is not too difficult to check that the following two assertions hold.

   (i) There exists a constant $C = C(c_1, C_2) > 0$ so that for every $A' \subset A$ and $B' \subset B$ that satisfy $|A'| + |B'| > n$:

$$\sum_{uv \in E, u \in A', v \in B'} p_{uv} \geq Cn.$$

   (ii) The random subgraph H_p of H almost surely does not contain a perfect matching.

   The validity of (i) can be checked directly; (ii) follows from the fact that with high probability not many more than $n/4$ edges of H_2 will survive in H_p and the edges of H_1 cannot contribute more than $2c_1 n < n/2$ edges to any matching.

Acknowledgement I would like to thank Dimitris Bertsimas for bringing the problem addressed here to my attention, Joel Spencer for helpful comments and Svante Janson for fruitful and illuminating discussions.

References

[1] N. Alon and J. H. Spencer, **The Probabilistic Method**, Wiley, 1991.

[2] B. Bollobás, **Random Graphs**, Academic Press, 1985.