Graph Partitioning Induced Phase Transitions

Gerald Paul,[1,*] Reuven Cohen,[1,2] Sameet Sreenivasan,[1,3] Shlomo Havlin,[1,4] and H. Eugene Stanley[1]

[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[2]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA*
[4]*Minerva Center and Department of Physics, Bar Ilan University, Ramat Gan 52900, Israel*

(Received 14 February 2007; published 10 September 2007)

We study the percolation properties of graph partitioning on random regular graphs with N vertices of degree k. Optimal graph partitioning is directly related to optimal attack and immunization of complex networks. We find that for any partitioning process (even if nonoptimal) that partitions the graph into essentially equal sized connected components (clusters), the system undergoes a percolation phase transition at $f = f_c = 1 - 2/k$ where f is the fraction of edges removed to partition the graph. For optimal partitioning, at the percolation threshold, we find $S \sim N^{0.4}$ where S is the size of the clusters and $\ell \sim N^{0.25}$ where ℓ is their diameter. Also, we find that S undergoes multiple nonpercolation transitions for $f < f_c$.

DOI: 10.1103/PhysRevLett.99.115701 PACS numbers: 64.60.Fr, 05.45.Df, 64.60.Ak

The graph partitioning problem deals with assigning vertices in a graph to different partitions such that no partition is greater than a given size. The optimal solution is one which minimizes the fraction of edges f that must be removed such that there are no edges between partitions [1].

Graph partitioning is of interest not only because of the large amount of previous research done [1–7] but also because optimal partitioning is equivalent to optimal attack or immunization of a complex network. That is, the percolation threshold f_c, at which global connectivity is lost, will be lower than that for any other type of attack or immunization and the measure of fragmentation F [8] for all values of f will be higher than for any other type of attack or immunization [9].

The study of optimal attack is relevant to both determining how to attack a network and how to best design a network. The attacker can compare the complexity of a proposed strategy to the complexity of an optimal attack and decide whether the incremental complexity of the optimal attack is justified. The study is also of value to the network designer; for any given network design under consideration, study of the optimal attack against the network tells the designer what the minimum cost will be to the potential attacker. The network designer can then ensure that a network is not over-engineered (i.e., designed for a network which far exceeds the capabilities of the attacker). Because immunization is equivalent to network attack, our work is applicable to the problem of how to effectively immunize populations against disease with minimal cost. This is an important issue; the number of deaths caused by major epidemics dwarfs the total number of deaths on all past battlefields [10].

Here we study graph partitioning from the standpoint of statistical physics. To make contact with percolation theory [11,12], we identify the number of edges removed as the control variable and study the inverse problem: Given that we are allowed to remove a fraction f of the edges from the graph, how can we partition the graph to minimize the size of the largest partition? We denote as S the size of the largest connected component (cluster) which results from the partitioning [13]. Then, S plays the role of order parameter and we are interested in the behavior of S as a function of f. We ask if there is a critical value f_c such that for $f < f_c$, $S \sim N$ while for $f > f_c$, S scales slower than $O(N)$. That is, does the graph undergo a percolation phase transition? If so, what is the percolation threshold f_c and what are the critical exponents associated with the phase transition.

We study random *k-regular* graphs, random graphs the vertices of which all have the same degree k. The graphs are constructed using the *configuration model* [14–16]. We study these graphs because of their intrinsic interest and because these graphs are examples of expander graphs which are extremely robust to node or edge removal [17,18]. They are therefore a good testbed for optimal graph partitioning.

We find that, in fact, a percolation transition does exist and we analytically determine f_c. We also estimate critical exponents associated with the transition. In addition, however, we find that for $f < f_c$ the graph undergoes a large number of first order transitions related to the partitioning process.

Percolation threshold.—The percolation threshold can be determined analytically as follows. In Refs. [19,20] it was argued that for a random graph having a degree distribution $P(k)$ to have a spanning cluster, a vertex j which is reached by following a link (from vertex i) on the giant cluster must have at least one other link, on average to allow the cluster to exist—otherwise the spanning cluster is fragmented. Thus at the critical point,

$$\langle k_i | i \leftrightarrow j \rangle \equiv \sum_{k_i} k_i P(k_i | i \leftrightarrow j) = 2 \quad (1)$$

where the angular brackets denote an ensemble average, k_i is the connectivity of node i, and $P(k_i|i \leftrightarrow j)$ is the conditional probability that node i has connectivity k_i, given that it is connected to node j.

We will show below that, for large N at f_c, all partitions are essentially the same size and that each partition consists of one cluster [21]. Then, to achieve Eq. (1) the average degree in each cluster must be 2 and p_c the fraction of edges which must be present is

$$p_c \equiv 1 - f_c = \frac{2}{k}. \quad (2)$$

This is to be compared to the random site or bond percolation threshold $p_c = 1/(k-1)$ [19].

We can gain insight into the structure of the spanning clusters by noting that for tree graphs with n vertices

$$\langle k \rangle = \frac{2(n-1)}{n} \quad (3)$$

which approaches 2 as $n \to \infty$. For finite graphs, however, to satisfy $\langle k \rangle = 2$, there must be on average one loop in each graph. Thus, at the percolation threshold, the clusters contain on average one loop. Our problem can be restated as follows: How do we partition a graph into the largest number of equal sized partitions each composed of one cluster with on average one loop per cluster? The larger the number of partitions (and thus the smaller the partition size), the closer the solution is to the optimal one. Different types of partitioning that maintain one cluster per partition will result in the same critical point but the scaling of the cluster size at the critical point may depend on the optimality of the partitioning.

Optimal partitioning.—We use the METIS graph partitioning program [6] which provides close to optimal graph partitioning. For the same random graph we run the program at least 100 times over the range of partition sizes in which we are interested. After each partitioning we identify the clusters in the graph, determine the size of the largest cluster, and note the number of edges needed to be removed for the partitioning. For each value of the number of edges, we maintain the minimum value of the size of largest cluster in the partitioning.

Figure 1(a) illustrates the behavior of $s \equiv S/N$ versus f for various values of k [22]. In what follows we will analyze the case $k = 3$ in depth; similar results are obtained for other values of k. In the inset in Fig. 1(a) for $k = 3$ and various values of N we plot s versus f. Below f_c the plots collapse indicating that here $S \sim N$. In the vicinity of and above f_c the plots no longer collapse, a manifestation of S scaling more slowly than N.

In Fig. 1(b), for $N = 10^6$, we plot $P(S)$ the distribution of cluster sizes S versus S at the threshold predicted by Eq. (2) $f_c = 1/3$. As expected, the distribution is very strongly peaked—almost all clusters are the same size; the inset shows the distribution of sizes of blobs which are discussed below.

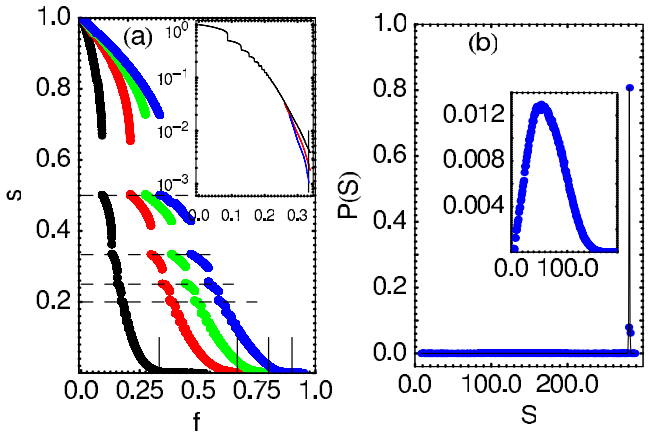

FIG. 1 (color online). (a) Normalized largest cluster size, $s \equiv S/N$, versus fraction of edges removed, f, for random regular graphs with $N = 10^4$ vertices of degree (from left to right) $k = 3$, 6, 10, and 20. Vertical lines at the x axis mark the predicted values of $f_c = 1 - 2/k$ from left to right for $k = 3$, 6, 10, and 20. Dashed horizontal lines at $s = 1/2$, $1/3$, $1/4$, and $1/5$ are the values of s for which the first few nonpercolation transitions take place. Inset: For (from top to bottom on right) $N = 10^4$, 3×10^4, and 10^5 and $k = 3$, s versus f. Data collapse until f is in the vicinity of $f_c = 1/3$ (indicated by vertical line). (b) For $N = 10^6$ and $k = 3$ at criticality, $P(S)$, the distribution of cluster sizes S. The inset is a plot of $P(S_B)$, the distribution of blob sizes, S_B, for $N = 10^6$ and $k = 3$.

In Fig. 2(a) we plot S_c the value of S at the percolation threshold versus N. The plot's slope is consistent with

$$S_c \sim N^x, \quad (4)$$

where $x \approx 0.4$. In Fig. 2(b), we plot S_c versus N for various values of f and see that the straightest plot is for $f_c = 1/3$, the predicted critical threshold.

In Fig. 2(a) we also plot ℓ the *diameter*, the maximum chemical distance between any two vertices of a cluster, of the critical clusters versus N. The slope of the plot is consistent with

$$\ell \sim N^z, \quad (5)$$

where $z \approx 0.25$. From Eqs. (4) and (5) we obtain

$$S_c \sim \ell^{d_l}, \quad (6)$$

where $d_l \equiv x/z \approx 1.6$. The exponent d_l is a measure of the compactness of the clusters: Clusters with $d_l = 1$ are essentially chains; higher values of d_l correspond to more dense structures.

Figure 3(b) is a representative critical cluster obtained from partitioning. Note the single loop required by Eq. (1) and its "stringy" structure, the manifestation of $d_l \approx 1.6$. In Fig. 3(a) we plot the distribution of the number of loops per cluster $P(n_{\text{loop}})$ and note that it is fairly narrow with the

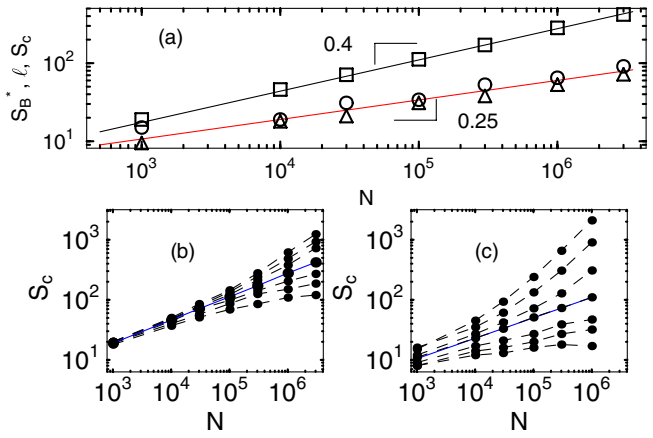

FIG. 2 (color online). (a) Largest cluster size at criticality, S_c (squares), diameter of largest cluster, ℓ (circles), and most probable blob size S_B^* (triangles), versus number of vertices N in graph. (b) For $k = 3$, largest cluster size for (from top to bottom) values of $f = 0.331$, 0.332, 0.333, 1/3 (solid line), 0.335, 0.337, and 0.34, versus number of vertices N in graph. The straightest plot is for $f = 1/3$, the predicted value of f_c. (c) Same as (b) for $k = 6$. For (from top to bottom) values of $f = 0.64$, 0.65, 0.66, 2/3 (solid line), 0.68, 0.69, and 0.70, versus number of vertices N in graph. The straightest plot is for $f = 2/3$, the predicted value of f_c.

most probable value being 1. Thus, not only is the average number of loops per cluster 1, but the most probable number is also 1.

We next determine the fractal dimension of the spanning clusters when the clusters are embedded in Euclidean space of dimension equal to the upper critical dimension. The upper critical dimension d_c is defined such that for dimension $d \geq d_c$, all critical exponents are unchanged. At or above the critical dimension the exponent $\tilde{\nu}$ is defined by [12]

$$r \sim \ell^{\tilde{\nu}}, \tag{7}$$

where r is Euclidean distance. At the percolation threshold, $\tilde{\nu}$ is expected to be 1/2, the same value as for a random walk (or for a network embedded in a very high dimensional lattice, such that spatial constraints are irrelevant) [12].

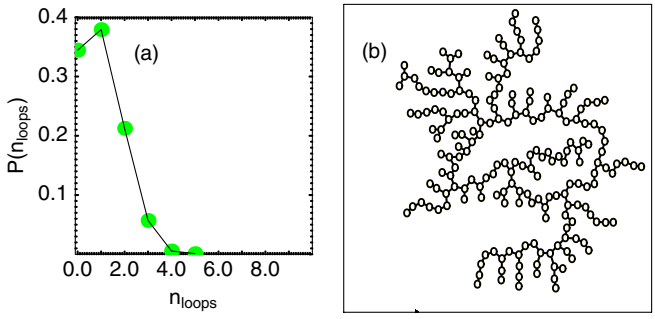

FIG. 3 (color online). (a) At criticality for $N = 10^5$ and $k = 3$, $P(n_{\text{loop}})$ distribution of number of loops per cluster, n_{loop}. (b) For $N = 10^5$ and $k = 3$, typical cluster at criticality containing 1 loop decorated by trees.

Using Eq. (6) with $d_l = 1.6$ and Eq. (7) with $\tilde{\nu} = 1/2$, we can determine the fractal dimension of the percolation clusters at criticality defined by $S_c \sim r^{d_f}$ to be

$$d_f = \frac{d_\ell}{\tilde{\nu}} \approx 3.2. \tag{8}$$

We now determine the upper critical dimension d_c. Using the fact that in Euclidean space $N \sim r^d$, we find $S_c \sim N^{d_f/d_c} \sim N^{0.4}$ and thus $d_c = 8$, which interestingly is the critical dimension for lattice animals and branched polymers [23,24].

We can learn more about the fractal structure of the spanning cluster at f_c by analyzing the 2-connected components (blobs) [25] within the spanning clusters. This is equivalent to analyzing the loops within the spanning clusters because the typical cluster contains 1 loop which is the 2-connected component in the cluster. In Fig. 2(a) we plot the most probable blob size (equivalent to the length of loops), S_B^*, versus N. The scaling is consistent with $S_B^* \sim N^{0.25}$ similar to the scaling of the diameter of the whole cluster. From this we infer that the diameter of the cluster is driven by the size of the loops.

Nonoptimal partitioning.—We find that for partitioning in which we ensure that each partition consists of one cluster but no attempt is made to minimize the number of edges between partitions, as predicted above, f_c in this case is also $1 - 2/k$ but at criticality $S \sim N^{1/2}$. That is, the clusters at criticality are larger than those at criticality for optimal partitioning. The argument that the exponent is exactly 1/2 is as follows: We ask how large a cluster must be to have on average one loop. Consider a cluster of size S. The total number of edges associated with vertices in the cluster is kS. Connectivity among vertices in the cluster is provided by $S - 1$ of the edges and others (also of order S) are either removed (connected to other partitions) or connected back to the cluster forming a loop. Because the graph is random and we partition randomly (subject to the constraint that the partitions consist of one cluster each), the probability that one of these edges is connected back to the cluster is

$$P_{\text{loop}} \sim S\frac{S}{N}. \tag{9}$$

Setting $P_{\text{loop}} = 1$ we find $S \sim N^{1/2}$.

Random partitioning.—Random partitioning is achieved by assigning vertices to partitions randomly and is equivalent to random site percolation [26], for which the well known result $f_c = 1 - 1/(k - 1)$ holds [19,20]. In contrast to the optimal and the nonoptimal partitioning considered above, for random partitioning, partitions contain clusters of all sizes (including very small ones). Equa-

tion (2) holds for the spanning cluster in each partition but does not hold for all clusters and f_c is therefore larger.

Nonpercolation transitions.—In Fig. 1(a), we see that the order parameter is discontinuous at values of $s = 1/2, 1/3, \ldots$, qualifying these points as first order phase transitions. However, these discontinuities, which occur where the number of partitions changes are not percolation transitions—the scaling of s with N does not change. The behavior at these transitions (and the general shape of the segments of the plots) can be understood as follows: Consider the region of the plot corresponding to two partitions ($1/2 < s < 1$) and assume we reduce the size of the larger partition (increasing the size of the smaller partition) by moving selected vertices one-by-one from the larger partition to the smaller partition [27]. Initially, the number of edges needed to be removed when we move a vertex is k—all edges adjacent to the moved vertex must be removed. As the size of the smaller partition increases, we can select a vertex requiring fewer of its edges to be removed because some of its edges already have ends in the smaller partition. At some point, the number of edges to the smaller partition of a vertex to be moved is equal to the number of the vertex's edges to the larger partition—thus, there is zero cost to the move [28]. This continues to be the case until the partitions are of equal size, resulting in the discontinuity.

Discussion.—Random regular graphs, due to their degree homogeneity, are the most difficult to attack or immunize. We have chosen to study these graphs to prove the feasibility and efficiency of our partitioning method for attack or immunization for this baseline class of network. Our optimal partitioning method suggests a new direction for studying and improving strategies for attack or immunization of many types of complex networks. It is of interest from an application standpoint to study also such networks as scale-free networks which represent many real world systems. Because our method is optimal we expect significant improvement over known attack or immunization strategies for these networks. We also believe that our method may be of value in finding network communities since it partitions the network into strongly connected components.

We thank ONR, the Israel Science Foundation, and the Dysonet Project for support. We have benefited from discussion of our results with Noga Alon. His analysis confirms our results for f_c and concludes that the component size is polylogarithmic at the transition point, implying that our estimates for the critical exponents may reflect the heuristic nature of the algorithm used. We thank Professor Alon for sharing his results with us.

*gerryp@bu.edu

[1] B. Kernighan and S. Lin, Bell Syst. Tech. J. **49**, 291 (1970).
[2] P. Buser, Discrete Appl. Math. **9**, 105 (1984).
[3] B. Bollobás, Comput. Math. Org. Theory **9**, 241 (1988).
[4] N. Alon, Comb. Probab. Comput. **6**, 145 (1997).
[5] S. L. Bezrukov, in *Graph Theory and Combinatorial Biology*, edited by L. Lovász *et al.* (Bolyai Society Mathematical Studies, Budapest, 1999), Vol. 7, p. 157.
[6] G. Karypis and V. Kumar, SIAM J. Sci. Comput. **20**, 359 (1998); METIS, a Software Package for Partitioning Unstructured Graphs, see http://www.cs.umn.edu/karypis/metis.
[7] S. Boettcher and A. G. Percus, Phys. Rev. E **64**, 026114 (2001).
[8] S. P. Borgatti, Comput. Math. Org. Theory **12**, 21 (2006); Y. Chen *et al.*, Phys. Rev. E **75**, 046107 (2007).
[9] P. Holme *et al.*, Phys. Rev. E **65**, 056109 (2002).
[10] R. M. Anderson and R. A. May, *Infectious Diseases in Humans* (Oxford University Press, Oxford, 1992).
[11] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992).
[12] A. Bunde and S. Havlin, *Fractals and Disordered Systems* (Springer, Berlin, 1995).
[13] One must distinguish between *partition* and *cluster*. A partition may contain one or more clusters.
[14] A. Bekessy, P. Bekessy, and J. Komlos, Stud. Sci. Math. Hung. **7**, 343 (1972).
[15] E. A. Bender and E. R. Canfield, J. Comb. Theory Ser. A **24**, 296 (1978).
[16] M. Molloy and B. Reed, Rand. Struct. Alg. **6**, 161 (1995); Comb. Probab. Comput.**7**, 295 (1998).
[17] P. Sarnak, Not. Am. Math. Soc. **51**, 762 (2004).
[18] L. Donetti *et al.*, J. Stat. Mech. (2006) P08007.
[19] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).
[20] R. Cohen, D. ben-Avraham, and S. Havlin, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, New York, 2002), Chap. 4.
[21] This is a reasonable assumption because an optimal algorithm will not needlessly remove edges in a partition which results in multiple clusters; the solution would not be optimal. However, one cannot, *a priori*, rule out optimal partitioning resulting in multiple clusters per partition.
[22] The quantity s is equivalent to the order parameter in percolation theory P_∞, the fraction of sites in the incipient infinite cluster.
[23] T. C. Lubensky and J. Isaacson, Phys. Rev. Lett. **41**, 829 (1978).
[24] S. Havlin *et al.*, Phys. Rev. Lett. **53**, 178 (1984).
[25] H. J. Herrmann and H. E. Stanley, Phys. Rev. Lett. **53**, 1121 (1984).
[26] Each partition is equivalent to a percolation system after vertices have been randomly removed from the system and is thus equivalent to random site percolation.
[27] This is not necessarily how optimal partitioning would occur but it is illustrative of the point we are making.
[28] Here we consider graphs with even k. For odd k, consider moving 2 vertices at a time.