Architecture of idiotypic networks: Percolation and scaling Behavior

Markus Brede and Ulrich Behn
Institut für Theoretische Physik, Universität Leipzig, Augustusplatz 10, D-04109 Leipzig, Germany
(Received 7 June 2000; revised manuscript received 12 January 2001; published 20 June 2001)

We investigate a model where idiotypes (characterizing *B* lymphocytes and antibodies of an immune system) and anti-idiotypes are represented by complementary bit strings of a given length *d* allowing for a number of mismatches (matching rules). In this model, the vertices of the hypercube in dimension *d* represent the potential repertoire of idiotypes. A random set of (with probability *p*) occupied vertices corresponds to the expressed repertoire of idiotypes at a given moment. Vertices of this set linked by the above matching rules build random clusters. We give a structural and statistical characterization of these clusters, or in other words of the architecture of the idiotypic network. Increasing the probability *p* one finds at a critical *p* a percolation transition where for the first time a large connected graph occurs with probability 1. Increasing *p* further, there is a second transition above which the repertoire is complete in the sense that any newly introduced idiotype finds a complementary anti-idiotype. We introduce structural characteristics such as the mass distribution and the fragmentation rate for random clusters, and determine the scaling behavior of the cluster size distribution near the percolation transition, including finite size corrections. We find that slightly above the percolation transition the large connected cluster (the central part of the idiotypic network) consists typically of one highly connected part and a number of weakly connected constituents and coexists with a number of small, isolated clusters. This is in accordance with the picture of a central and a peripheral part of the idiotypic network and gives some support to idealized architectures of the central part used in recent dynamical mean field models.

DOI: 10.1103/PhysRevE.64.011908 PACS number(s): 87.18.−h, 64.60.Ak, 05.10.Ln, 02.70.Rr

I. INTRODUCTION

B lymphocytes carry on their surface highly specific receptors, so-called antibodies. If these receptors detect complementary structures, the lymphocyte is stimulated to proliferate and after several generations and differentation into plasma cells secretes antibodies of the same specificity. *B* lymphocytes and antibodies of a given specificity are said to have a certain idiotype. Complementary structures to an idiotype are antigens or other anti-idiotypic antibodies. Between *B* lymphocytes of different idiotypes thus emerges a functional network of mutual stimulation and inhibition, the idiotypic network [1]. The idiotypic network is supposed to contribute at least partially to the functionality of the immune system, e.g., to immunological memory or suppression of autoreactive clones. Although quantitative data are very hard to access by experiment there are some recent observations that underline the importance of idiotypic interactions [2,3].

New idiotypes are produced in the bone marrow or due to hypermutation during the proliferation of stimulated lymphocytes which introduces a random metadynamics of the repertoire. At a given moment the random network has a certain architecture. The aim of this paper is to give a statistical description of this architecture. Knowledge of the typical architecture of the idiotypic network is crucial for describing the population dynamics of the interacting *B* lymphocytes and antibodies, cf. [4,5], which is, however, not the subject of this paper.

Derived from hypotheses of theoretical immunology (cf. [1,6]; for a recent review see [7]) we present a statistical analysis of bit-string based networks, which shows that this approach is well suited to reproducing reasonable network structures. In particular, our considerations show that realistic network topologies can be conceived as an extension of first approaches to that problem which assumed a Bethe-lattice structure. Suggestions of how loops should be added to such structures have been made previously [8]. For this, our work provides a very natural access.

Generally, idiotypic networks are supposed to realize a tradeoff between two basic requirements: they should contain a great number of small isolated components, but on the other hand still be able to respond to arbitrary antigens, which means being complete. Small components are thought necessary to store information about previously encountered antigens [7,9,10]. The existence of such components obviously demands a low connectivity of the network. With completeness, on the other hand, it is assumed that of a great number of antibodies each is able to detect many different types of antigen. Hence the network connectivity should not be too low.

With regard to the underlying biological problem the one great–many small cluster situation merits special attention. Theoretical immunologists suppose that the idiotypic network consists of a large number of small clusters. On the other hand, as a consequence of relatively high connectivities, a large component should also be contained within the idiotypic network, which is denoted as its central part. This great cluster could play an important role in the control of autoreactive clones [6,11–13].

In the following the bit-string model of [14] will be explained briefly. Basically, antibodies are identified with bit chains of a given length *d*. Thus, there are 2^d different antibody types. The set of all conceivable antibodies, i.e., the potential repertoire, is then represented by $\{i=(i_1,\ldots,i_d), i_j \in \{0,1\}\}$. Estimating the probable size of networks that can be complete in the above sense shows that $d \approx 32-36$ should be a good value for realistic models [15].

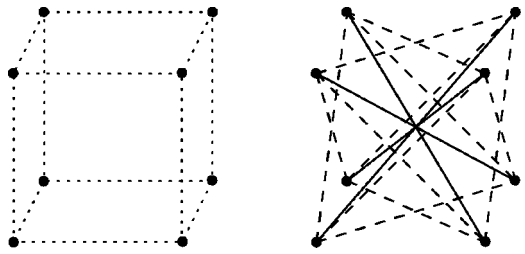

FIG. 1. Comparison of a hypercube (left side) with the one-mismatch base graph (right side) that has the same set of vertices ($d=3$). Solid lines connect perfectly complementary vertices; dashed lines mean one-mismatch links.

Antibodies recognize each other if they have complementary structures, i.e., if they are represented by perfectly complementary bit strings. If there are small deviations from the exact structural complementarity a matching is still possible, although with lower affinity. This is described by so-called matching rules.

For example, we imagine that antibodies react if the respective structures are complementary except in one small region. Using the language of our model the corresponding bit strings should be inverse except for one bit that belongs to the noncomplementary region. We call this kind of rule a one-mismatch rule. Two-mismatch rules, which allow reactions between antibodies that are complementary up to two nonmatching regions, are accordingly defined by connecting bit strings that are complementary with two exceptions allowed. Naturally, the multitude of possible rules is not exhausted by mismatch rules. Rules that express matching of mutually shifted antibody parts could be conceived as well. For the sake of simplicity this work is confined to those matching rules that are associated with reactions of highest affinity, i.e., inversion and one- and two-mismatch rules.

Mathematically, the set of all antibody types (represented by bit strings or vertices of a hypercube) together with all possible reactions between them defines a graph. Thus, we want to call

$$G_d^1=(\{v=(v_1,\ldots,v_d),v_j \in\{0,1\}\},\{v \text{ connected with w if } v_i=\overline{w_j} \text{ for all } i,j=1,\ldots,d \text{ except one position at the most}\}) \quad (1)$$

the one-mismatch base graph and

$$G_d^2=(\{v=(v_1,\ldots,v_d),v_j \in\{0,1\}\},\{v \text{ connected with w if } v_i=\overline{w_j} \text{ for all } i,j=1,\ldots,d \text{ except two positions at the most}\}) \quad (2)$$

the two-mismatch base graph. To obtain a better impression of one-mismatch graphs they can be compared to hypercubes of the same dimension. Looking, e.g., on hypercubes of dimension $d=3$, edges of one-mismatch graphs are represented by all space and side diagonals. This, together with a picture of a one-mismatch graph in dimension 4 is shown in Figs. 1 and 2, respectively.

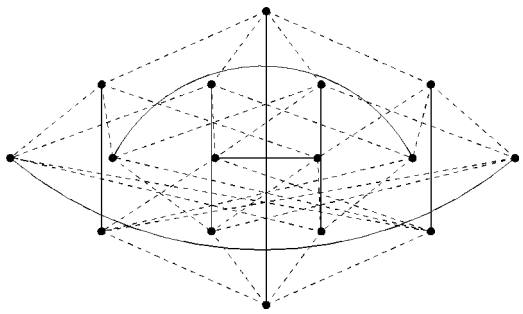

FIG. 2. One-mismatch base graph in $d=4$. Same notation as in Fig. 1.

The actual repertoire, i.e., the set of all types of idiotype that really exist within the body at a given time, is a subset of the potential repertoire. There are profound reasons to believe that the actual repertoire should be distributed randomly within the potential repertoire [7]. A way of realizing this is to choose antibodies with a certain probability p. Then the idiotypic network is represented by the random graph that is composed of all occupied vertices (antibodies) and their connections (possible reactions). Clearly, this is a site percolation problem [16].

Percolation problems on standard lattices are far from being a new field of research. Percolation on one-mismatch graphs, however, has some major differences from ordinary percolation.

Percolation is always connected to an abrupt change of some system property (the standard property is the existence of a connecting path from the upper to the lower boundary on two-dimensional lattices) if the occupation probability passes a certain value, the percolation threshold. This transition becomes sharp in the limit of infinite systems.

On lattices there is no question of how to increase the system size to infinity. However, sequences of one-mismatch graphs exhibit relations between small and large systems that are essentially different from similar characteristics of lattices, which are created by the multiplication of a unit cell. Understanding of these distinctions comes from studying the structure of fully occupied one-mismatch graphs (base graphs) which will be performed in the next section.

In this paper we investigate the architecture of functional networks built by constituents of randomly generated characteristics interacting with complementary constituents. Our motivation, the formulation of the problem, and the interpretation of the results are all in the context of modeling the immune system using the language and methods of statistical physics. However, our results are essentially independent of the specific immunological interpretation and could be of broader interest. The situation of interacting randomly generated complementary constituents is quite general. For example, think of chemical reactions (origin of life), ecosystems, or social networks.

Generally, the concept of the paper is as follows. In Sec. II topological properties of base graphs are studied. Section III gives an analysis of the underlying percolation problem, which then leads, together with applications of random graph theory from Sec. IV and a study of intrinsic structures of great clusters in Sec. V, to conclusions about parameter re-

gimes, in which bit-string model induced random graphs resemble idiotypic networks. In the Appendix another method of calculating thresholds via the renormalization of small cells is discussed.

II. STRUCTURE OF THE BASE GRAPH FOR THE ONE-MISMATCH RULE

The above definition of matching rules allows an easy calculation of distances between vertices of G_d^1. Let $i,j \in G_d^1$ be vertices, d_H their Hamming distance (i.e., the number of different bits between them), and

$$d_G(i,j) = \inf_{\substack{\text{paths } w \text{ in } G_d^1 \\ \text{connecting } i \text{ and } j}} l(w), \tag{3}$$

where $l(w)$ denotes the length of a path w, a metric. Then it holds that if

$$d_H < d/2, \qquad d_G = \begin{cases} d_H & \text{for } d_H \text{ even} \\ d_H + 1 & \text{for } d_H \text{ odd}, \end{cases}$$

$$d_H = d/2, \qquad d_G = d/2, \quad \text{for } d \text{ even}, \tag{4}$$

$$d_H > d/2, \qquad d_G = \begin{cases} d - d_H + 1 & \text{for } d - d_H \text{ even} \\ d - d_H & \text{for } d - d_H \text{ odd}, \end{cases}$$

which shows that the maximum distance $D_{\max}(d)$ between vertices is roughly one-half of that on ordinary hypercubes. This is essentially due to the inversion rule. Furthermore, it should be noted that $D_{\max}(d)$ does not change when increasing the dimension by 1 from odd ($d=2n+1$) to even ($d=2n+2$) values.

To obtain additional information about the toplogy we distinguish vertices corresponding to their distance s to an arbitrarily chosen origin, e.g., $v=0=(0,\ldots,0)$, whose choice is arbitrary due to the specific construction rule of base graphs. We denote the set of vertices with distance $d_G(v,i)$ from v by

$$E_s = \{ i \in G_d^1 | d_G(v,i) = s \}. \tag{5}$$

Applying Eq. (4) it is possible to compute the number of vertices $|E_s|$ belonging to E_s (mass distribution) by simple combinatorics to

$$|E_s| = \begin{cases} \begin{pmatrix} d \\ d_G - 1 \end{pmatrix} & \text{if } d_G = \dfrac{d+1}{2} \\ \\ \begin{pmatrix} d+1 \\ d_G \end{pmatrix} & \text{if } d_G \leqslant \dfrac{d}{2} \end{cases}. \tag{6}$$

A visualization of Eq. (6) in Fig. 3 shows a major difference between odd and even dimensions d. Generally, $|E_s|$ grows monotonically with increasing distance. For odd dimensions, however, the number of vertices with maximal distance $D_{\max}$ from v is substantially smaller than that with distance $D_{\max}-1$.

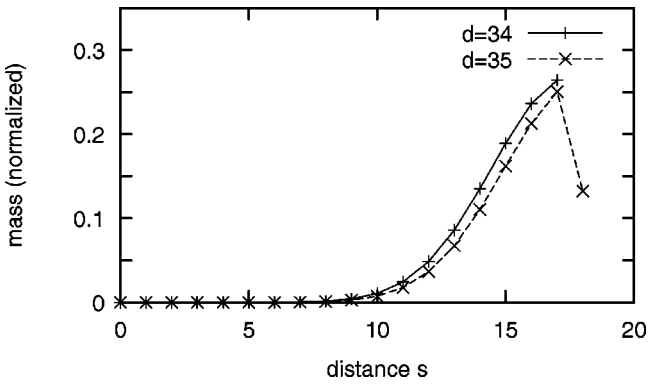

FIG. 3. Normalized mass in the dimensionless distance s from an arbitrary origin in even ($d=34$) and odd ($d=35$) dimensions. Points have been connected to guide the eyes.

Differences in the overall structure of graphs in odd and even dimensions d become even more obvious if the number of links connecting vertices of the same set E_s is considered. Using Eq. (4), a somewhat lengthy but straightforward calculation shows that connected vertices in the same distance s from the origin v occur in

$$\begin{aligned} &d \text{ even only for} \quad s = D_{\max} = d/2, \\ &d \text{ odd only for} \quad s = D_{\max} = (d+1)/2 \ \text{ or} \\ &\qquad\qquad s = D_{\max} - 1. \end{aligned} \tag{7}$$

Analogously, the number of links that connect vertices of E_s with vertices belonging to E_{s+1} may be computed to $d+1-s$ for $s < D_{\max}-1$.

As a consequence there must exist loops of even as well as of odd length. Moreover, all short loops with length less than d must be even (since only loops comprising links between vertices in the same distance from any loop element can be of an odd length).

An important property of standard systems in percolation theory is that the smaller system is always contained within the larger one. For one-mismatch graphs, however, this is not possible. Let $d_1 < d_2$ be dimensions of $G_{d_1}^1$ and $G_{d_2}^1$, respectively. If $G_{d_1}^1$ were contained in $G_{d_2}^1$ it would follow directly that $G_{d_2}^1$ should have uneven loops with length less than d_1. This gives a contradiction since $G_{d_2}^1$ contains no odd loops smaller than d_2, whereas $d_1 < d_2$ by assumption.

Summarizing these results we state that percolation on one-mismatch graphs differs from standard percolation. A crucial point in this distinction is the way global system properties change if the system size approaches infinity. Nevertheless, methods of percolation theory can be applied to this kind of problem. Further, by choosing an appropriate majority rule renormalization group procedures can be extended to one-mismatch graphs (see Appendix A).

To the best of our present knowledge percolation problems on graphs with similar properties have not been dealt with. Even percolation on regular hypercubes has had relatively little attention in the physical literature [17].

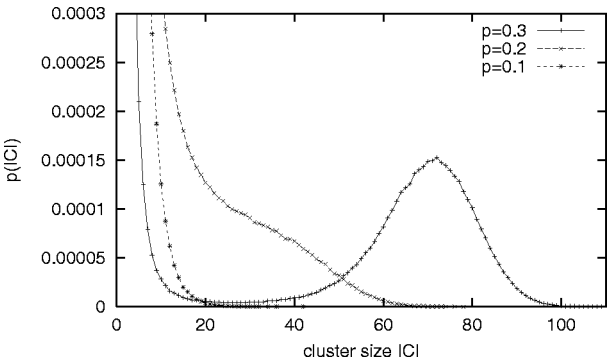

FIG. 4. Cluster size distribution depending on the occupation probability $p=0.1,0.2,0.3$ for base graphs in eight dimensions. The histogram was taken over 10^5 configurations.

III. CLUSTER SIZE DISTRIBUTION

Having collected some simple properties of the underlying base graphs we consider now vertices of those graphs occupied with a given probability p. The set of all occupied vertices Γ together with all bonds, that connect two vertices belonging to Γ forms a random graph. In terms of percolation theory a maximal set of connected vertices is called a cluster.

What is the probability that an arbitrarily chosen vertex belongs to a cluster of size $|C|$? This question has already been addressed in [9,10] where three major regimes of the system have been identified (see also the numerical results in Fig. 4).

(i) The first typical situation arises for small values of the occupation probability p. Then only small clusters (whose individual size does not make up a finite fraction of the whole system) are expected to appear.

(ii) On increasing p and approaching infinite dimensions, a sharp transition (in the following denoted as the percolation transition) to a one great cluster–many small clusters regime was found. At the percolation transition some characteristics obey scaling laws. These and an approximation to compute percolation thresholds will be dealt with in Sec. IV.

The one great–many small clusters situation deserves special attention (cf. Sec. I). The great cluster could play an important role in the control of autoreactive clones. To fulfill that purpose it is believed to have a certain internal structure (see [12,13,18]; for experimental data see [19]), which will be discussed in more detail in Sec. VI. Thus, as stated already in [9,10], the simple one-parametric bit-chain model exhibits an interesting similarity to the idiotypic network.

(iii) Finally, if the occupation probability is further increased, a state will be reached where random graphs consist of one connected component only . Relying on some general results of the theory of random graphs it can be shown that this indeed marks a second (in the limit of infinite systems) sharp transition which we call the denseness transition. In Sec. V these considerations are recalled and discussed with reference to the completeness of the immune system.

Exactly the same two thresholds were considered earlier for a different class of random graphs [20,21,26] and more recently in the context of networks of RNA secondary structures [22,23].

IV. SCALING LAWS AND BETHE APPROXIMATION

A common method of analyzing percolation problems is the introduction of perimeter polynomials $D_{|C|}$. Using the standard notation we define

$$p(|C|)=p^{|C|}\sum_{S_{\mathrm{free}}(C)}(1-p)^{|S_{\mathrm{free}}(C)|}=p^{|C|}D_{|C|}(1-p), \quad (8)$$

where $p(|C|)$ denotes the probability that a vertex on the base graph belongs to a cluster of size $|C|$. $S_{\mathrm{free}}(C)$ means the free surface, i.e., the set of all unoccupied vertices of the base graph that are adjacent to vertices of C.

Generally, there are few problems that allow the explicit calculation of $D_{|C|}$ for arbitrary cluster size $|C|$. Notwithstanding, it is always possible to compute $D_{|C|}$ for small values of $|C|$. This provides a basis for the application of series expansion techniques. We investigate the structure of small clusters up to size $|C|=5$ on one-mismatch graphs. As a result for large base graphs ($d>5$) we find the following perimeter polynomials for one-mismatch graphs:

$$D_1(q)=q^{d+1}, \quad (9)$$

$$D_2(q)=(d+1)q^{2d}, \quad (10)$$

$$D_3(q)=\frac{3}{2}(d+1)dq^{3d-2}, \quad (11)$$

$$D_4(q)=q^{4d-4}\left(2(d-1)d(d+1)+\frac{d(d+1)}{2}+\frac{2}{3q}(d-1)d(d+1)\right), \quad (12)$$

$$D_5(q)=\frac{5}{2}q^{5d-6}(d-1)d(d+1)\left(d-2+\frac{1}{q}(d-1)+\frac{1}{q^2}+\frac{1}{12q^3}(d-2)\right), \quad (13)$$

where $q=1-p$.

In the following we show that in the limit of high dimensions random graphs on one-mismatch graphs for small values of p are very similar to random graphs on Bethe lattices with the same coordination number. Arguments of this kind have already been applied to percolation on hypercubic lattices [24,25] making use of the fact that loops in random graphs on sparsely occupied high dimensional lattices are infrequent. Indeed, the first terms of a high dimension expansion on a hypercubic lattice of dimension d are given by the exact values of the percolation threshold for the Bethe lattice with coordination number d.

A closer look at the perimeter polynomials (9)–(13) supports the hypothesis that the situation is quite similar for one-mismatch-graphs. For example, the contribution from

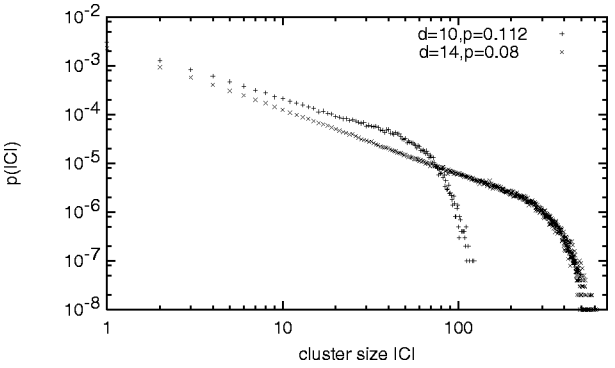

FIG. 5. Cluster size distribution on G_{10}^1 and G_{14}^1 at the percolation thresholds. The simulation data are plotted on doubly logarithmic scales.

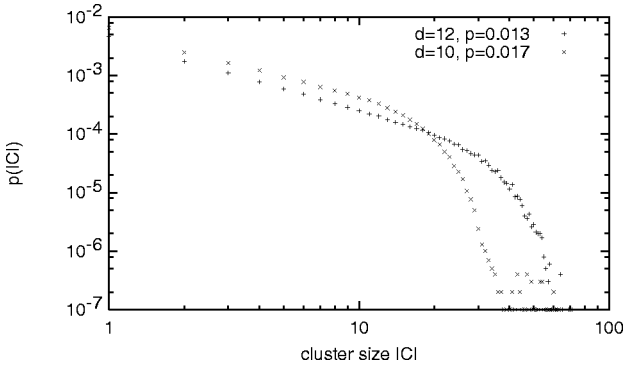

FIG. 6. Illustration of finite size scaling laws for the two-mismatch graphs G_{10}^2 and G_{12}^2. Data sets are plotted on doubly logarithmic scales.

loops to $p(4)=p^4D_4(q)$ is given by $p^{\text{loop}}(4)=p^4q^{4d-4}d(d+1)/2$. Obviously, the relative weight of this term vanishes in the high dimension limit. Additional support comes from the observation that the one-mismatch base-graph G_d^1 contains hypercubes up to dimension $d-1$. To elucidate this property we consider the following example in $d=4$. We construct a hypercube of dimension 3, applying the one-mismatch rule allowing for a mismatch in one of the, last three bits. Starting at the origin (0,0,0,0) this yields (1,1,1,0), (1,0,1,1), (1,1,0,1). Iteration gives the new vertices (0,1,0,1), (0,0,1,1), (0,1,1,0), and finally (1,0,0,0). These eight vertices are connected by matching rules like those of a three-dimensional hypercube. Then follow the arguments of [24,25].

Consequently the percolation threshold for one-mismatch graphs can be approximated to

$$p_c^{(1)}=1/d, \tag{14}$$

whereas for two-mismatch graphs

$$p_c^{(2)}=1/[d+d(d+1)/2]. \tag{15}$$

Corrections to Eqs. (14) and (15) are of order $O(d^{-2})$ and $O(d^{-3})$, respectively. From considerations of Sec. I we find that the number of links connecting vertices in E_s with vertices in E_{s+1} decreases with increasing distance s. Thus, we conclude that the corrections to $p_c^{(1)}$ must be positive.

Typically for percolation problems, certain system characteristics obey scaling laws at the percolation threshold. This kind of law reflects the statistical self-similarity of clusters on different length scales in that parameter regime. Figure 5 displays simulation results for the cluster size distributions obtained for one-mismatch graphs of dimensions 10 and 14. Both data sets can be well described by the finite size scaling ansatz

$$p(|C|)=|C|^{-\tau}F(|C|/|C|^*), \tag{16}$$

where the function $F(x)$ is nearly a constant for $x\ll 1$ and will be more rapidly declining than any power law for arguments $x\gg 1$. Equation (16) should apply for clusters that are neither too small nor too large. Due to the limited size of the systems investigated, very large clusters behave in a different way from clusters of ''standard'' size. The transition between the laws applying to these separate cases is marked by $|C|^*$, which in turn depends on the extent of the system.

Analogous properties can be observed for base graphs defined by other matching rules. Figure 6 illustrates scaling behavior of random graphs on two-mismatch base graphs for values of $p=p_c^{(2)}$ [cf. Eq. (15)] As a natural consequence of adding supplementary edges distances between vertices become smaller than in one-mismatch graphs. This gives an explanation for the fact that typical sizes $|C|^*$ are substantially smaller than those examined on one-mismatch graphs with the same choice of d.

Furthermore, it can be seen in Fig. 6 that the scaling law (16) applies well even to very small clusters. It is useful to define a cluster size dependent exponent $\tau_{|C|}$ by

$$\tau_{|C|}=-\frac{\ln[p(|C|+1)/p(|C|)]}{\ln[(|C|+1)/|C|]}. \tag{17}$$

Evaluating the perimeter polynomials (9)–(13), values of $\tau_{|C|}$ for small $|C|$ can be derived. Thus we obtain a sequence $\{\tau_{|C|}\}$ which should approach the true value of τ for large $|C|$.

As a matter of fact Eq. (16) changes to $p(|C|)\sim|C|^{-\tau}$ in the limit of infinite systems. Performing the limits $d\to\infty$ in Eqs. (9)–(13) and (17), we observe that $\tau_{|C|}^{\infty}$ obeys

$$\tau_{|C|}^{\infty}=\frac{1}{\ln(1+1/|C|)}-|C|+1, \tag{18}$$

which by comparison of Eq. (17) with Eq. (18) and solving the recursion relation implies the law

$$p(|C|)=\frac{e^{-|C|}|C|^{|C|-2}}{|C|!} \tag{19}$$

for the cluster size distribution. Equation (18) leads to $\tau=3/2$ for large clusters. The value $\tau=3/2$ thus computed is in accord with the exact result on the Bethe lattice.

Since we investigate a model for a biological system our main interest is devoted to large, yet finite systems. In the sequence $\{\tau_{|C|}\}$ our best approximation for the real value of

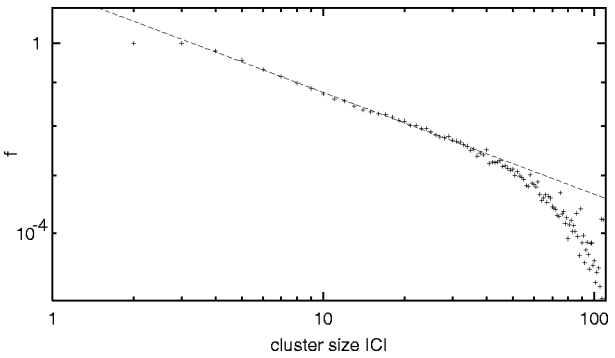

FIG. 7. Simulation results for the fragmentation rate f on G_{10}^{1} depending on the cluster size $|C|$ at the numerically determined $p_c \approx 0.112$. Data are plotted on doubly logarithmic scales. Fitting a power law (dashed line) yields $\lambda \approx 0.12$ to be compared with $\lambda \approx 0.11$ from Eq. (24).

τ is τ_4. A comparison between values of τ computed from $\tau \approx \tau_4$ and exponents τ obtained by evaluating numerical data for small systems suggests that the approximation involved becomes rapidly more accurate if the system size is increased. Hence in $d=32$ we rely on τ_4 and find $\tau \approx 1.5$, which gives a result that also supports the previous assumption.

Subsequently it was our aim to find a quantity that gives an overall estimation of deviations between random graphs on one-mismatch graphs from those on Bethe lattices of equal coordination number. For this purpose it appears appropriate to investigate the results of an edge elimination procedure which computes the number of bonds belonging to loops. Similar procedures have very recently been applied to lattices [27].

The above aim is achieved by individually removing every edge of every cluster and calculating the number of connected components of the resulting graph. An edge is said to be fractioning if two components are obtained by cutting this edge. Clearly, only edges that belong to loops are not fractioning.

Then the ratio of fractioning bonds $f(C)$ to the overall number of bonds $b(C)$ of a cluster C, $f(C)/b(C)$, is some indication of the importance of loops within the structure of C. Distinguishing clusters according to their size $|C|$ we computed the mean fractioning ratio

$$f_{|C|} = \langle f(C')/b(C') \rangle_{|C'|=|C|}. \tag{20}$$

Except for very small clusters with trivial structure we expect a finite size scaling law

$$f_{|C|} = |C|^{-\lambda} \hat{F}(|C|/|\hat{C}^*|) \tag{21}$$

for the fragmentation rate f at the percolation threshold, the validity of which is illustrated in Fig. 7.

Similar to $\tau_{|C|}$, a cluster size dependent exponent $\lambda_{|C|}$ may be defined. The only 4- and 5-clusters containing nonfragmenting bonds are the 4-loop (no fractioning bond) and the 4-loop with a tail (1/5 of all bonds fractioning) which give the contributions $p^{\text{4-loop}}(4)$ and $p^{\text{tailed 4-loop}}(5) = 5/2 p^5 q^{5d-7}(d-1)d(d+1)$ to $p(4)$ and $p(5)$, respectively. Making use of Eqs. (12) and (13) we then find

$$f_4 = \frac{p(4) - p^{\text{4-loop}}(4)}{p(4)} = \frac{1}{1 + 1/4(d-1)[1 + 1/(3q)]} \tag{22}$$

and

$$f_5 = \frac{p(4) - p^{\text{tailed 4-loop}}(5)}{p(4)} + \frac{1}{5}\frac{p^{\text{tailed 4-loop}}(5)}{p(4)} = 1 - \frac{4}{5}\frac{1}{1 + 1/q + (d-2)[1 + q + 1/(12q^2)]}. \tag{23}$$

Finally, inserting $p \approx p_c^{(1)} = d^{-1} + O(d^{-2})$ a rough estimate for λ is $\lambda_4 = (\ln f_5 - \ln f_4)/\ln(5/4)$, which gives

$$\lambda_4(d) \approx \frac{1}{\ln(5/4)} \left(\frac{393}{2 \times 10^3} d^{-1} + \frac{4\,029\,120}{8 \times 10^6} d^{-2} + O(d^{-3}) \right). \tag{24}$$

For $d=10$ we have $\lambda \approx 0.11$ in very good agreement with the numerically obtained value 0.12 (see Fig. 7). For $d=32$ we find $\lambda \approx 0.03$, already very close to $\lambda = 0$ which holds for Bethe-lattices. Thus the value $\lambda > 0$ measured on finite dimensional random graphs on one-mismatch base graphs—which is caused by a small number of loops—should quantify the deviation from random graphs on Bethe lattices. It appears that the occupation probability at the percolation threshold is still small enough to apply the Bethe approximation.

Yet, random graphs at p_c are not exactly Bethe-like and contain a certain fraction of loops. Otherwise it would be impossible to distinguish subclusters according to their connectivity within any considerable connected component. This, however, is likely to be necessary, to explain the role of the central part of idiotypic networks properly [13].

Further investigations concerning the structure of great clusters will be made in Sec. V.

V. DENSITY THRESHOLD AND THE COMPLETENESS OF THE IDIOTYPIC NETWORK

Hitherto, our main interest was devoted to the question of the occupation probability at which a great cluster starts to exist. However if p, is further increased it is also imaginable that a situation occurs where the whole system consists of one great cluster only. We call this the question about the connectivity property of random graphs.

In the theory of random graphs some general results that address related problems have been derived for so-called sequences of configuration spaces [22,23] and earlier for a different general class of random graphs in [20,21]. It can be shown that the sequence of one-mismatch base graphs $\{G_d^1\}$ fulfills all requirements of these configuration spaces [32]; because of its rather technical nature the proof will be omitted here.

General results of [22,23] can be applied here showing

that for infinite systems there exists a threshold

$$p_{\mathrm{conn}} = 1 - \lim_{d\to\infty} |G_d^1|^{-1/\gamma_d} \tag{25}$$

(γ_d being the coordination number) with the property that almost all random graphs are connected for $p > p_{\mathrm{conn}}$ while the set of all connected random graphs has measure zero for $p < p_{\mathrm{conn}}$. Moreover, for sequences of configuration spaces it holds that both the connectivity and the density threshold, i.e. the threshold for the property that there is no nonoccupied site without occupied neighbors on the base graph, coincide. Using Eq. (25) we find

$$p_{\mathrm{conn}}^{(1)} = p_{\mathrm{dense}}^{(1)} = 1/2 \tag{26}$$

in the case of one-mismatch graphs and

$$p_{\mathrm{conn}}^{(2)} = p_{\mathrm{dense}}^{(2)} = 0 \tag{27}$$

for two-mismatch graphs.

The density of random graphs in our model can be directly translated into biological terms. Since antibodies and antigens are represented by the same sets of bit chains, the property that there is no free site without an occupied neighbor means that every antigen is sure to encounter a complementary antibody. Hence the idiotypic network is able to respond to any antigen. This is the completeness axiom [7] for the immune system.

Nevertheless, it is somewhat difficult to reconcile the demands for denseness of random graphs and the occurrence of small clusters at the same time. However, it seems unlikely that completeness should be understood in this strict way. Rather it appears to be a better solution to consider the fact that evolution has most likely driven the idiotypic network into such an arrangement that it is able to respond only to variations of actually existing antigens. Thus we argue that the completeness of the idiotypic network does not exactly match the density of random graphs, but requires only the probability for the density of the corresponding random graph to be somewhat below 1. So random graphs can still comprise small clusters and be complete.

Another paradox arises if two-mismatch graphs are considered. For those the connectivity and percolation thresholds $p_c^{(2)}$ and $p_{\mathrm{conn}}^{(2)}$ fall together. How then could small clusters and a separate large component coexist if all random graphs are connected? Yet, for the case of finite systems it is clear that $p_c^{(2)}(d) < p_{\mathrm{conn}}^{(2)}(d)$, i.e., a great cluster has to be formed first, before it can encompass its competing small rivals.

Consequently, for the case of two-mismatch graphs also the biologically interesting regime is well defined. For finite systems there is a range of occupation probabilities $p_c^{(2)}(d) < p < p_{\mathrm{conn}}^{(2)}(d)$, where all requirements are met. Because both probabilities $p_c^{(2)}(d)$ and $p_{\mathrm{conn}}^{(2)}(d)$ are converging to zero in the limit $d\to\infty$ it follows that the extent of this range $|p_c^{(2)}(d) - p_{\mathrm{conn}}^{(2)}(d)|$ will also become very small for large systems.

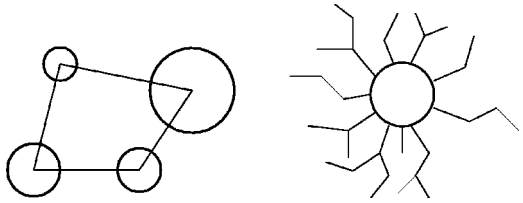

FIG. 8. Visualization of different cluster structures. The circles symbolize strongly connected subclusters (heaps). Left hand side: the cluster consists of several weakly connected heaps. Right hand side: the cluster contains one heap only.

In the next section our interest will be shifted toward the intrinsic structure of great clusters. Further conclusions about biologically relevant parameter regimes can then be drawn.

VI. INTRINSIC STRUCTURE OF GREAT CLUSTERS

Summarizing the results for percolation on mismatch graphs from Secs. I–IV, two phase transitions have been found to occur. The ranges below p_c (since there is no great cluster) and above p_{conn} (since there is only one great cluster) are of no interest for the biological background of the model. The one great cluster–many small clusters situation between the two, however, is likely to fulfill some requirements for idiotypic networks (see Sec. II). In this section further investigations into the network structure within this parameter regime will be made.

Insights into the structure of a great cluster can be obtained from the mass distribution of this cluster $M_c(s) = \{v \in C | d(v,c) = s\}$, i.e., the information on how many vertices have a certain distance s from a starting point c. Due to the equality of all starting points c we define

$$M(s) = \frac{1}{|C|} \sum_{c\in C} M_c(s) \tag{28}$$

and consider the mean value of $M(s)$ calculated over all clusters whose size exceeds a minimum value $0.5 \times p \times 2^d$. For the metrics $d(\cdot,\cdot)$ there are two distinct choices, viz., metrics defined by Eq. (3) allowing paths in $G = G_d^1$, i.e., $d(\cdot,\cdot) = d_G$, and such restricting paths to C itself, i.e., $d(\cdot,\cdot) = d_C$. Since $d_C > d_G$ we confine our investigations to $d(\cdot,\cdot) = d_C$, which provides a better ''resolution.'' In the following we will discuss some typical cluster compositions to obtain a survey of possible conclusions from mass distributions about cluster structures.

Three different situations can be imagined [for (i) and (iii) see Fig. 8].

(i) A cluster could consist of some loosely bound high connectivity regions (heaps), whose vertices are distinguished by a great number of connections with each other. Other vertices are bound with relatively few links. Clearly, strong connectivity within heaps means that vertices belonging to them have nearly the same distance from all other vertices. Thus, the presence of many heaps should result in a great number of local extrema. On the other hand, loosely bound vertices will smooth the mass distribution, i.e., reduce the sharpness of extrema.

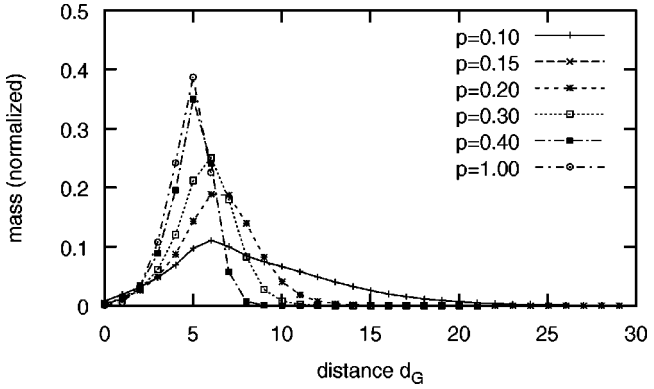

FIG. 9. Simulation results for the normalized mass distribution on G_9^1 for varying probabilities $p=0.1,0.15,0.2,0.3,0.4$ and the results of Eq. (6) for $p=1.0$. In order to make results for different p's comparable on the same scale masses have been divided by the cluster size $|C|$. Points have been connected to guide the eyes.

(ii) A cluster could contain no distinguishing parts at all. From Eq. (6) we know that up to $s=D_{\max}$ the number of vertices is increasing with increasing distance s. For small occupation probabilities $p<p_{\text{conn}}$ (for which the whole base graph has not been covered yet) we expect a compromise between the declining probability that a vertex of that cluster has distance s from the origin v and the increasing number of vertices with increasing s. Consequently, the mass distribution should exhibit one maximum. For large $p>p_{\text{conn}}$ mass distributions can simply be derived by multiplying Eq. (6) by p.

(iii) As a special case of (i) a cluster could be made of one heap and a certain fraction of loosely bound vertices. Then up to two maxima, caused by the heaped vertex concentration and the competing tendencies [see (ii)], respectively, are likely to occur. The sharpness of both maxima will depend on the fraction of loosely bound vertices. Thus, in the case of a large proportion of those vertices both extrema could be smoothed to just one broad maximum.

Figure 9 shows simulation results for the normalized mass distribution for different values of p. Clearly, all distributions are marked by only one maximum, whose sharpness increases with increasing values of p. In the vicinity of the density threshold clusters are already stretched over the whole extent of the base graph. Hence, for $p=0.4$, slightly below $p_{\text{conn}}=0.5$, the curve of the mass distribution looks similar to the exactly known distribution for the fully occupied base-graph given by Eq. (6); see also Fig. 3.

More interestingly, for small p's in the vicinity of the percolation threshold relatively broad maxima occur, which could be an indication of cluster structures as described in (iii).

To prove the validity of this hypothesis we have applied the edge-elimination algorithm (see Sec. II) to great clusters. As a slight extension of the described procedure all fractioning edges are removed, thus splitting a cluster C into the sequence of its doubly connected components, or heaps in the above sense, $(C_1,\ldots,C_t)$. Obviously, vertices belonging to doubly connected components are distinguished by their many connections in comparison to other vertices.

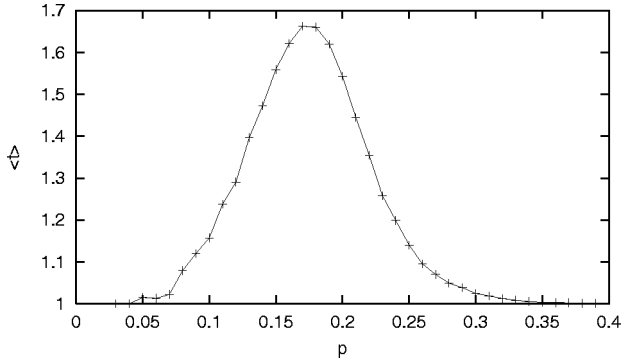

FIG. 10. Mean value $\langle t\rangle$ of the resulting large fragments after application of the edge-elimination procedure for G_9^1 depending on the occupation probability p.

Thus, the number of such components t should allow one to distinguish between the situations (i)–(iii).

Figure 10 shows simulation data for the mean value $\langle t\rangle$ of the resulting nontrivial parts after edge elimination depending on p. The distribution is marked by one maximum, which is again a consequence of two competing tendencies. For small values of p clusters are generally not doubly connected, but increasing p leads to a larger proportion of vertices that belong to loops. On the other hand, there is a tendency for loops to get connected by loops, i.e., separate doubly connected cluster components grow together. Thus, for large p almost every cluster will consist of one doubly connected component only.

Results of Fig. 10 show also that the maximum is reached slightly above the percolation threshold p_c. Then the number of doubly connected parts t rapidly declines until it asymptotically approaches $t=1$ for $p\to 1$. We argue that this behavior is caused by one great doubly connected component which occurs first for some $p\gtrsim p_c$ and then gradually incorporates all other doubly connected parts. In principle, these results do also apply to two-mismatch graphs (see Fig. 11).

Comparing this with the scenario described above of cluster structures, we can thus state that there is a range of values of $p\gtrsim p_c$ where (iii) applies, i.e., clusters are made of one

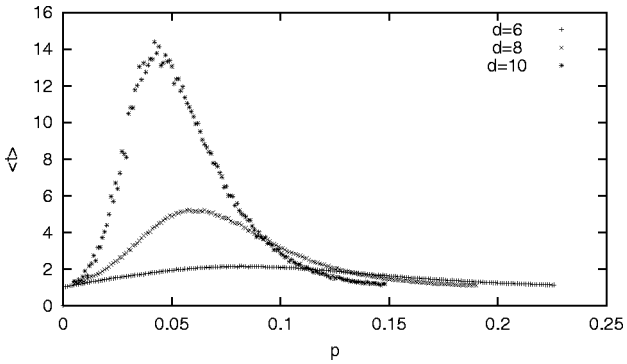

FIG. 11. Mean value $\langle t\rangle$ of the resulting great fragments for the two-mismatch graphs G_6^2, G_8^2 and G_{10}^2 depending on p. Except for a shift toward smaller values of p, the results are similar to those on one-mismatch graphs.

great doubly connected component (including several small ones) and a set of other loosely bound vertices. Accordingly, within this range of p we define two subsets of great clusters C, namely, the great doubly connected component $B=C^{(2)}$ and the complementary set $P=C-B$.

As already discussed in Sec. I, the central part of idiotypic networks (corresponding to great clusters C) should contain strongly and weakly connected distinguished subsets. From the previous analysis it becomes clear that there is a choice of the only parameter p where bit-string models can exactly reproduce such a situation.

VII. CONCLUSIONS

We investigated statistical properties of a bit-string-based model for idiotypic networks and compared typical architectures of the network thus defined with axioms and hypotheses for idiotypic networks from theoretical biology. Before introducing randomness we undertook an analysis of the underlying base graphs to show a major difference from standard percolation problems based on the way the size of the system is increased.

Subsequently the expressed antibody repertoire was identified with graphs, created by randomly occupying a matching rule defined base structure (base graphs). Concepts of percolation theory were applied in order to determine the percolation threshold.

The immune system is a very large, yet finite system. Consequently, finite size corrections have to be taken in account. Series expansion techniques allowed the calculation of two critical exponents (for finite systems and in the limit $d\rightarrow\infty$) that characterize the scaling behavior. A regime of values for the parameter $p>p_c$ has been found where random graphs consist of many small clusters and one great connected component. For choices $p\gtrsim p_c$ our model reproduces the peripheral/central part concept for idiotypic networks.

Translation of the notation of the completeness of idiotypic networks into the language of graph theory allowed the determination of an upper threshold p_{dense} for the occurrence of the many small clusters–one great cluster situation. Furthermore, relying on general results of the theory of random graphs (for so-called sequences of configuration spaces) we calculated the (in this case coincident) thresholds for the density (p_{dense}) and connectivity (p_{conn}) properties.

Furthermore, we developed techniques to obtain additional information about the structure of great clusters. Analyzing random graphs for $p\gtrsim p_c$ great clusters can be decomposed into two subsets of vertices with different binding properties, namely, groups of doubly connected vertices (backbone) and loosely linked vertices (peripheral part of the great cluster).

Thus, bit-string models are suited to describe a hierarchy of connectivity levels, which really existing idiotypic networks are also expected to exhibit. Our results support to some extent idealized architectures used in a mean-field type model to describe the dynamics of the central part of the immune system [13]. Unlike other model approaches [8,28–31] for topologies of idiotypic networks, our simple few parametric bit-string model produces a nontrivial seemingly realistic network topology, without assuming *a priori* distinguished vertex groups.

APPENDIX A: RENORMALIZATION GROUP APPROACH

In this Appendix we present a method to approximate the connectivity threshold for one-mismatch graphs. Extension of the ideas of renormalization group theory could also be applicable to more complicated cases that do not allow an exact treatment.

Our approach is based on the idea of the renormalization of small cells [33] which employs the self-similarity of the system on different length scales at the percolation threshold. A condition to apply this procedure is that an appropriate grouping of sites on the original lattice leads to a renormalized lattice with the same symmetry properties as the original one. Here, treating not a real space lattice but functional networks, we adapt the idea of the renormalization group theory in the following way. We find a transformation $\mathcal{R}$ which, by grouping of vertices, leads from the one-mismatch base graph G^1_{d+2} in dimension $d+2$ to a graph $\mathcal{R}(G^1_{d+2})$ that is equivalent to the base graph G^1_d in dimension d, $\mathcal{R}(G^1_{d+2})\simeq G^1_d$, thus allowing a systematic reduction of the degrees of freedom.

If a threshold property appropriately defined for finite systems, namely, $p_{\mathrm{conn}}(d)$, converges to a certain value for $d\rightarrow\infty$, differences between p_{d} and $p_{\mathrm{d-2}}$ must be small and will disappear in the limit of infinite systems. Consequently, if a grouping of vertices to supervertices on a graph G yields a renormalized graph $\mathcal{R}(G)$ of the same type, percolation thresholds on both graphs will be the same. Thus, we replace the term symmetry (as it is applied to lattices) by equivalency of graphs.

We encode an arbitrary vertex of the base graph G^1_{d+2} by (A,b_0,b_1) where A is a bit chain of length d and b_0,b_1 are single bits. Every vertex on G^1_{d+2} belongs to a unique 4-loop $\{(A,b_0,b_1),(\overline{A},\overline{b_0},b_1),(A,\overline{b_0},\overline{b_1}),(\overline{A},b_0,\overline{b_1})\}$ connected by one-mismatch links. The renormalization $\mathcal{R}$ replaces this 4-loop by the super vertex (A) on $\mathcal{R}(G^1_{d+2})$ [of course the choice of $(\overline{A})$ leads to identical results]. If two vertices A and B are connected by an inversion (one-mismatch) link on G^1_d the vertices of the corresponding 4-loops on G^1_{d+2} are connected by four inversion (one-mismatch) links, too see Fig. 12.

We thus define a renormalized graph $\mathcal{R}(G^1_{d+2})$ composed of the vertices and edges explained above. It follows directly from our construction that $\mathcal{R}(G^1_{d+2})$ is equivalent to G^1_d.

To describe random graphs we apply the following majority rule: a supervertex on G^1_d is considered as occupied if at least two connected vertices of the corresponding 4-loop on G^1_{d+2} are occupied, i.e., this 4-loop is said to percolate. On the basis of this rule we obtain

$$p'=4p^2(1-p)^2+4p^3(1-p)+p^4, \tag{A1}$$

where p' denotes the (renormalized) probability that vertices of (G) are occupied if vertices on G were chosen with probability p.

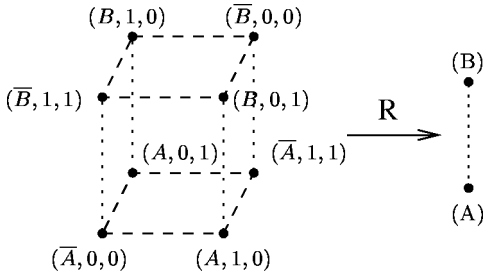

FIG. 12. Illustration of the renormalization procedure applied to a pair of connected 4-loops on G^1_{d+2} leading to a pair of linked supervertices on G^1_d. If $d_H(B,\overline{A})=0$ the dotted edges correspond to inversion links, if $d_H(B,\overline{A})=1$ to one-mismatch links.

Calculating fixed points of Eq. (A1) we obtain $p^*=0$, $(3\pm\sqrt{5})/2$, and 1. Thus, as the only unstable solution in $[0,1]$ we have $p^*=(3-\sqrt{5})/2$ as a first approximation for the connectivity threshold.

Here one could pose the question whether only 4-loops are suited as a renormalization cell. Indeed, it is also possible to summarize successive renormalization steps in to a single large one, and ordinary hypercubes of even dimension $k<d$ form suitable cells as well. On the other hand, condensing odd dimensional hypercubes to supervertices does not lead to renormalized graphs $\mathcal{R}(G)$ that are equivalent to other base graphs of the sequence $\{G^1_d\}$. Altogether, this seems to be an effect of differences between even- and odd-dimensional base graphs (see Sec. II).

Of course, renormalizing small cells entails some approximation. Grouping together vertices and applying a majority rule to occupy the renormalized vertices, situations can arise where clusters on the original lattice are cut or new clusters are formed [33].

There are different approaches to improve the approximations involved. One possibility suggested in [34], which leads to exact results in the limit of very large cells, is to summarize some elementary cells to one large cell of size z. This large cell will then be occupied if all elementary cells are occupied and connected, i.e., are said to percolate. Since ''renormalization faults'' are essentially due to cell surface effects improvements produced by the above method emanate from the declining surface-volume ratio of large cells. For one-mismatch graphs it holds, however, that the surface size of a cell $s(z)$ depends logarithmically on z, viz. $s(z)\approx d+1-\log_2 z$, leading to only slight improvements with increasing z.

Using larger elementary cells consisting of two coupled 4-loops we obtain

$$(p')^2=p^8+8p^7q+24p^6q^2+32p^5q^3+12p^4q^4, \quad \text{(A2)}$$

and using four coupled 4-loops yields

$$(p')^4=p^{16}+16p^{15}q+112p^{14}q^2+448p^{13}q^3+1120p^{12}q^4+1792p^{11}q^5+1776p^{10}q^6+1008p^9q^7+180p^8q^8 \quad \text{(A3)}$$

where $q=1-p$. As fixed points $p'=p$ of Eqs. (A2) and (A3) we determined numerically $p\approx 0.41$ and $p\approx 0.39$, respectively. The difference of both values from $p_{\mathrm{conn}}=0.5$ may be a consequence of the above mentioned slow convergence.

More interestingly, our renormalization procedure of replacing hypercubes of dimension k by supervertices is applicable without modifications to ordinary hypercubes as well. This gives an additional argument that the connectivity thresholds of both sequences of graphs are equal, which can also be verified by evaluating Eq. (25).

[1] N.K. Jerne, Ann. Inst. Pasteur Immunol. **125C**, 373 (1974).
[2] P.J. Ruiz *et al.*, Nat. Med. **6**, 710 (1998).
[3] C.A. Bona, Proc. Soc. Exp. Biol. Med. **213**, 32 (1998).
[4] R. J. de Boer, A. U. Neumann, A. S. Perelson, L. A. Segel, and G. Weissbuch, *Mathematics Applied to Biology and Medicine* (Wuerz Publishing, Winnipeg, 1993), pp. 243–258.
[5] A.S. Perelson, *Cell to Cell Signalling: From Experiments to Theoretical Models* (Academic Press, New York 1989), pp. 261–271.
[6] A. Coutinho, Immunol. Rev. **110**, 63 (1989).
[7] A.S. Perelson and G. Weisbuch, Rev. Mod. Phys. **69**, 1219 (1997).
[8] A. Neumann and G. Weisbuch, Bull. Math. Biol. **54**, 699 (1992).
[9] K. Lippert and U. Behn, Annu. Rev. Comput. Phys. **5**, 287 (1997).
[10] K. Lippert, Ph. D. dissertation, Universität Leipzig, [1994].
[11] I.R. Cohen, Immunol. Today **12**, 490 (1992).
[12] F.J. Varela and A. Coutinho, Immunol. Today **5**, 159 (1991).
[13] B. Sulzer, J.L. van Hemmen, and U. Behn, Bull. Math. Biol. **56**, 1009 (1994).
[14] J.D. Farmer, N.H. Packard, and A.S. Perelson, Physica D **22**, 187 (1986).
[15] A.S. Perelson and G.F. Oster, J. Theor. Biol. **81**, 645 (1979).
[16] D. Stauffer and A. Aharony, *Perkolationstheorie: Eine Einführung* (VCH, Weinheim, 1995).
[17] I.A. Campbell, J.-M. Flesselles, R. Jullien, and R. Botet, Phys. Rev. B **37**, 3825 (1988).
[18] J. Stewart and F.J. Varela, Immunol. Rev. **110**, 37 (1989).
[19] F. J. Kearney, M. Vakil, and N. Nicholson, in *Evolution and Vertebrate Immunity: The Antigen Receptor and MHC Gene Families*, edited by G. Kelsoe and D. Schulze (University of Texas Press, Austin, TX, 1987), pp.175-190.
[20] P. Erdös and A. Rényil, Publ. Math. Debrecen **6**, 290 (1959).
[21] E. M. Palmer, *Graphical Evolution* (Wiley, New York, 1985).
[22] C. Reidys, Ph.D. thesis, Friedrich-Schiller-Universität Jena, 1995.
[23] C. Reidys, P.F. Stadler, and P. Schuster, Bull. Math. Biol. **59**, 339 (1997).

[24] B. Kahng, Phys. Rev. A **43**, 1791 (1991).
[25] D.S. Gaunt, M.F. Sykes, and H. Ruskin, J. Phys. A **9**, 1899 (1976).
[26] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[27] M. Cheon, M. Heo, I. Chang, and D. Stauffer, Phys. Rev. E **59**, 4733 (1999).
[28] G.R. Weisbuch, R. De Boer, and A.S. Perelson, J. Theor. Biol. **146**, 483 (1990).
[29] A. Neumann and G. Weisbuch, Bull. Math. Biol. **54**, 21 (1992).
[30] R.W. Anderson, A. Neumann, and S. Perelson, Bull. Math. Biol. **55**, 1091 (1993).
[31] P.H. Richter, Eur. J. Immunol. **5**, 350 (1975).
[32] M. Brede, MS thesis, Universität Leipzig, 1999.
[33] P.J. Reynolds, H.E. Stanley, and W. Klein, Phys. Rev. Lett. **10**, 167 (1977).
[34] P.J. Reynolds, H.E. Stanley, and W. Klein, Phys. Rev. B **21**, 1223 (1980).