Network Topology of a Potential Energy Landscape: A Static Scale-Free Network

Jonathan P. K. Doye*

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW,United Kingdom

(Received 23 January 2002; published 20 May 2002)

Here we analyze the topology of the network formed by the minima and transition states on the potential energy landscape of small clusters. We find that this network has both a small-world and scale-free character. In contrast to other scale-free networks, where the topology results from the dynamics of the network growth, the potential energy landscape is a static entity. Therefore, a fundamentally different organizing principle underlies this behavior: The potential energy landscape is highly heterogeneous with the low-energy minima having large basins of attraction and acting as the highly connected hubs in the network.

DOI: 10.1103/PhysRevLett.88.238701 PACS numbers: 89.75.Hc, 31.50.–x, 61.46.+w

Energy landscapes have been at the forefront of many of the recent theoretical developments in our understanding of biomolecules [1], clusters [2,3] and the glass transition [4]. For example, this research has provided important new insights into how proteins fold [5]and the origin of the unusual properties of supercooled liquids, such as the distinction between "strong" and "fragile" liquids [6,7]. There has also been a surge of interest in modeling complex systems as networks [8], inspired by Watts and Strogatz's discovery that many networks behave as "small worlds" [9]. Intriguingly, a diverse range of such networks, e.g., the World Wide Web [10], the internet [11], scientific collaboration [12], and citation [13] networks, and biochemical networks [14,15], also have a "scale-free" topology, where the distribution of the number of connections to each node, the degree, follows a power law. This topology results from the dynamics of the network growth in these systems [16]. Here we draw these two strands of research together by applying the techniques of network analysis to probe the global structure of potential energy landscapes of clusters.

The potential energy landscape is a multidimensional surface representing the dependence of the potential energy on the positions of all the atoms of the system. For a system with many atoms the landscape will have a complex topography with higher-dimensional analogs of mountain ranges, valleys, and passes. Although the potential energy landscape determines the system's structure, thermodynamics, and dynamics, the nature of this relationship is complex. A particularly successful means of elucidating this relationship is the inherent structure approach of Stillinger and Weber [17], in which the landscape is divided into basins of attraction surrounding each minimum (see Fig. 1). This partition provides a natural way to describe the dynamics of the system, because, except at high temperature, the system spends most of the time vibrating in the well surrounding a minimum and only occasionally hops to a different well by passing over a transition state. The interbasin dynamics can then be represented as a walk on a network whose nodes correspond to the minima and where edges link those minimum which are directly connected by a transition state. Figure 1 provides an illustration of such an inherent structure network (ISN) for a two-dimensional energy surface.

The ISN should provide the starting point for an energy landscape view of the global dynamics of a system. Indeed, as is increasingly being done, it is relatively easy to calculate the dynamics from this network using a master equation approach [3,18,19]. However, fundamental questions about the topology of the ISN have received little attention. By contrast the global topography of energy landscapes has been the focus of much research [1,2,5]. To give one example, this emphasis has revealed that when a landscape is like a "funnel" [5] the system is guided towards the global minimum, be it the native state of a protein [5], an ordered nanoparticle [20], or a bulk crystal. However, the topological aspects of this idea remain open despite their importance: if the average number

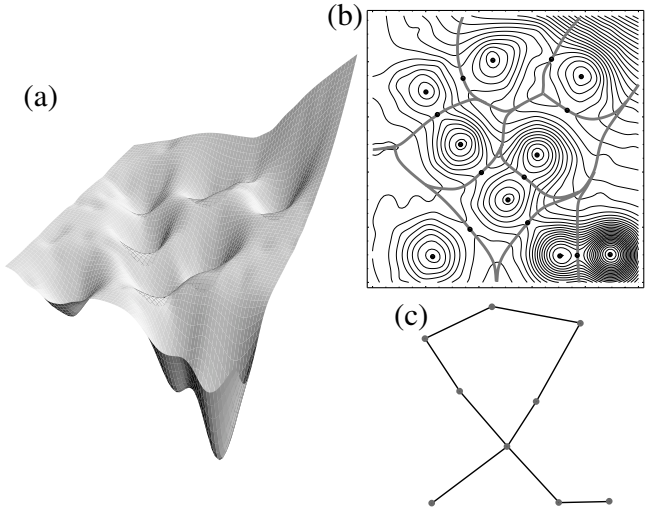

FIG. 1. (a) A model two-dimensional energy surface. (b) A contour plot of this surface illustrating the inherent structure partition of the configuration space into basins of attraction surrounding minima. The basin boundaries are represented by the thick lines, and the minima and transition states by points. (c) The resulting representation of the landscape as a network.

of steps to reach the global minimum from an arbitrary starting minimum scales unfavorably with size, the location of this structure would become significantly hindered at large size.

To characterize the topology of the ISN we study small clusters for which we are able to locate nearly all the minima and transition states on the potential energy landscape [3,21]. The atoms of the cluster interact with a Lennard-Jones potential, which provides a reasonable description for rare gas clusters. The numbers of minima and transition states are expected to increase roughly as $N_{\min} \approx e^{\alpha N}$ and $N_{\rm ts} \approx N e^{\alpha N}$, respectively [22,23], where N is the number of atoms in the cluster. Therefore, the largest network that we are able to consider is for a 14-atom cluster for which we have located 4196 minima and 87 219 transition states

Small-world networks have characteristics typical of both random graphs and lattices. The average separation between nodes scales logarithmically with network size, while the network is highly clustered, i.e., any two neighbors of a node are also likely to be connected. From Fig. 2 it is clear that the ISNs for the clusters show both these features and so are small worlds. The clustering is unsurprising given that the connections between basins on a potential energy landscape are based on adjacency in configuration space [24], but to interpret Fig. 2(a) properly we must take into account the increase in both the dimension of configuration space and the average degree $\langle k \rangle$ as the size of the network increases.

For example, for a hypercubic lattice with a constant number of lattice points L along each edge and dimension $3N$, the number of lattice points $N_{\rm latt}$ scales exponentially with N and the average number of steps between any two lattice points is $N(L+1) = (L+1)\log N_{\rm latt}/3\log L$. By contrast, if the network behaves as a random graph, the average separation should scale as $\log N_{\min}/\log\langle k \rangle \propto \log N_{\min}/\log(\log N_{\min})$ because $\langle k \rangle \propto N_{\rm ts}/N_{\min} \propto N$. The sublogarithmic scaling suggested by Fig. 2(a) points to the latter scenario. This result is somewhat surprising. In Watts and Strogatz's small-world model the random-graph character results from the introduction of random links into the lattice, which can potentially connect up distant nodes, but there is no obvious equivalent of the random links on the potential energy landscapes.

If we now examine the distributions for the numbers of connections for each node we find that as the size of the cluster increases a clear power-law tail develops, which has a universal form independent of the cluster size (Fig. 3). The exponent of the power law, $\gamma = 2.78$, is similar to other scale-free networks [25]. The cause of the random-graph like scaling of the average separation is thus the scale-free topology of the ISN. The network is extremely heterogeneous with a few hubs that have a very large number of connections, but with the majority of nodes only connected to a relatively small number of other minima.

This is a particularly surprising result because all other scale-free networks are dynamic in origin. They grow and change over time, be it on an almost continuous basis as in the WWW or on evolutionary time scales in the case of biochemical networks. Even the recently introduced deterministic scale-free networks are based on an iterative growth procedure [26,27]. Furthermore, models of network growth [16,25]and studies on the time evolution of real networks [28,29]strongly suggest that the heterogeneity at the heart of the scale-free topology develops as a result of new nodes preferentially linking to those nodes which have many connections, be they much-cited papers or popular web sites. However, the network associated with a potential energy landscape is static. It is simply determined by the form of the interatomic interactions and the number of atoms in the cluster.

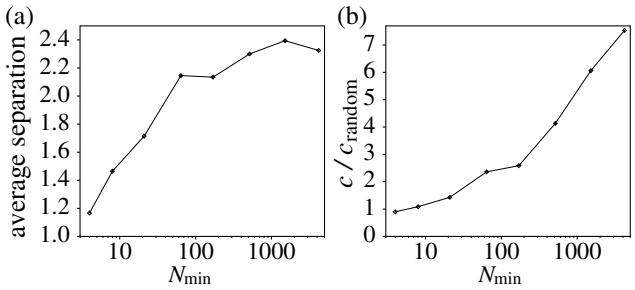

FIG. 2. (a) The dependence of the average separation between nodes (in steps) on the size of the network, $N_{\min}$. (b) The size dependence of the clustering coefficient, c, compared to that for a random graph, where c is the fraction of the pairs of nodes with a common neighbor that are also connected [9]. The data points are for Lennard-Jones clusters with from 7 to 14 atoms.

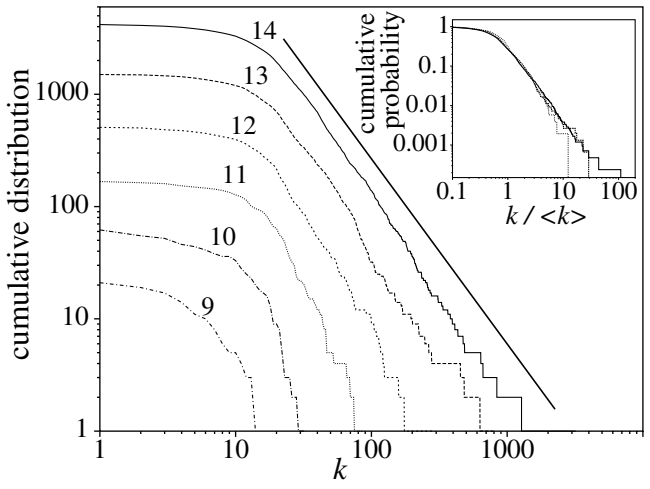

FIG. 3. The cumulative distribution for the number of nodes that have more than k connections. The curves correspond to clusters of different sizes, as labeled. An additional straight line with slope $-(\gamma - 1)$, where $\gamma = 2.78$, has been plotted to emphasise the power law tail. In the inset the cumulative probability distribution for the 12-, 13-, and 14-atom clusters is plotted against k normalized by its average value $\langle k \rangle$ to bring out the universal form of the distribution.

The source of the heterogeneity in the ISNs is apparent from Fig. 4, where we see that the degree of a node increases somewhat faster than exponentially as the energy of the minimum decreases [30]. The low-energy minima act as the hubs in the network. Thus, for the 14-atom cluster, 76% of the nodes in the network are connected to the global minimum. To measure the extent of the catchment basin around a minimum, we can calculate the distance in configuration space to all the transition states connected to a minimum. For LJ_{14} we find that this distance is 2.7 times larger for the global minimum than for the surrounding minima. When the multidimensionality of configuration space is taken into account this result suggests that the hyperarea of this catchment basin is many orders of magnitude larger than the average. Similarly, it has been previously found that on average the basin area falls off approximately exponentially with the energy of a minimum [31].

These results show that the global topology and topography of the potential energy landscape are intimately connected because the deep minima have very large catchment basins which are connected to lots of smaller basins around their edges. By contrast, if a potential energy landscape were flat and all basins of attraction had the same area the scale-free topology would be lost. For example, an investigation of the network topology of the configuration space of a *noninteracting* lattice polymer [32] found the connectivity distribution to be a Gaussian [33].

This contrasting example naturally leads one to ask how general is the topology that we have found for the Lennard-Jones clusters. Although such a question can be definitively answered only by similar analyses for a variety of systems, there is nothing "special" about the Lennard-Jones potential and so there is no reason why similar behavior should not be seen for other materials, as long as there are no constraints present that would prevent the formation of the high degrees associated with the hubs. A polymer provides an example of the latter, because the connectivity of the chain limits the number of transition states that can surround a minimum. For example, a similar analysis for Lennard-Jones polymers [34]did not find a power-law tail to the degree distribution (although it was still longer than exponential) [35]. There is no equivalent of many of the transition states for the equivalent Lennard-Jones cluster because they involve the breaking of the polymer chain.

The scale-free topology of the ISN is potentially good news for global optimization, the task of locating the global minimum. Even though the number of minima increases exponentially with the size of the system [22], the average number of steps in the shortest path to the global minimum grows sublinearly with system size. Of course, finding this path is not necessarily easy. Our calculations of the shortest paths required information on the global structure of the potential energy landscape, whereas a global optimization algorithm usually takes a step based on only local information.

Some path finding strategies to efficiently navigate scale-free networks have been suggested that make use of the fact that most of the shortest paths pass through the highly connected hubs [36,37]. In our case the clear link between the topology and topography of the potential energy landscape provides an additional advantageous strategy. A downhill step to a lower-energy minimum is likely to take one to a minimum that is more connected and closer to the global minimum. How well obeyed the latter correlation is, depends upon the global topography of the potential energy landscape and is a good indicator of the difficulty of global optimization. Thus, when the landscape is like a single funnel, global optimization algorithms can achieve near to the ideal scaling. For example, the basin-hopping algorithm can locate the global minimum of the 55-atom Lennard-Jones cluster after on average less than 150 minimizations when started from a random configuration, even though there are an estimated 10^{21} minima [31]. The increasing number of links as the energy decreases evident in Fig. 4 further adds to the efficacy of a funnel in guiding the system towards its bottom, and provides clear evidence of the convergence of pathways into the hub at the funnel bottom that is often postulated. By contrast, when an energy landscape has multiple funnels and there is a tendency to enter a funnel that leads the system to a low-energy minimum that is far from the global minimum, global optimization can be very difficult.

The topology of the ISN will of course significantly affect the dynamics. This connection can be probed for very small systems where the network can be completely characterized and the inherent structure dynamics obtained by a master equation approach. However, this approach is not practical for the system sizes that are of most

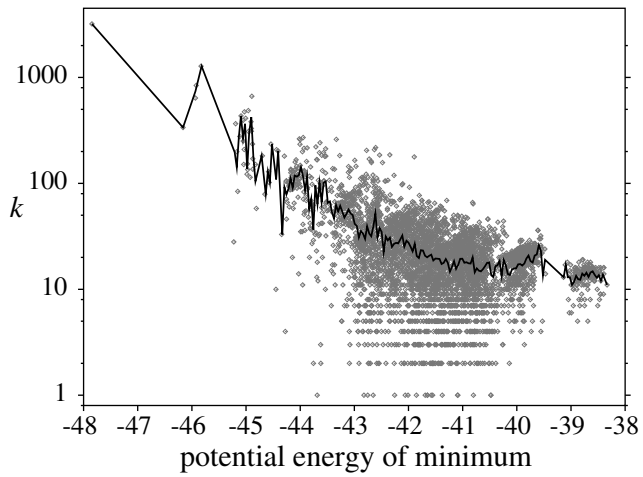

FIG. 4. The dependence of the degree of a node on the potential energy of the corresponding minimum for the 14-atom cluster. The data points are for each individual minimum and the solid line is a binned average.

interest. Therefore, models of protein folding and the glass transition usually have to assume a simplified topology for the interstate dynamics [38,39], or relate the dynamics to static quantities through phenomenological equations, such as the Adam-Gibbs equation which relates the relaxation time in supercooled liquids to the configurational entropy [40]. To fully unlock the potential insights from the inherent structure view of the dynamics, a means of statistically modeling the network topology from a partial characterization of the potential energy landscape is therefore needed. Our results could significantly advance this goal.

The author is grateful to Emmanuel College, Cambridge and the Royal Society for financial support.

*Electronic address jpkd1@cam.ac.uk

[1] C. L. Brooks III, J. N. Onuchic, and D. J. Wales, Science **293**, 612 (2001).
[2] D. J. Wales, M. A. Miller, and T. R. Walsh, Nature (London) **394**, 758 (1998).
[3] D. J. Wales *et al.,* Adv. Chem. Phys. **115**, 1 (2000).
[4] P. G. Debenedetti and F. H. Stillinger, Nature (London) **410**, 259 (2001).
[5] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct. Funct. Genet. **21**, 167 (1995).
[6] S. Sastry, Nature (London) **409**, 164 (2001).
[7] I. Saika-Voivod, P. H. Poole, and F. Sciortino, Nature (London) **412**, 514 (2001).
[8] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[9] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[10] R. Albert, H. Jeong, and A. L. Barabási, Nature (London) **401**, 130 (1999).
[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[12] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[13] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[14] H. Jeong *et al.,* Nature (London) **407**, 651 (2000).
[15] H. Jeong, S. Mason, A. L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[16] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[17] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[18] R. S. Berry and R. Breitengraser-Kunz, Phys. Rev. Lett. **74**, 3951 (1995).
[19] M. Cieplak, M. Henkel, J. Karbowski, and J. R. Banavar, Phys. Rev. Lett. **80**, 3654 (1998).
[20] K. D. Ball *et al.,* Science **271**, 963 (1996).
[21] C. J. Tsai and K. D. Jordan, J. Phys. Chem. **97**, 11 227 (1993).
[22] F. H. Stillinger, Phys. Rev. E **59**, 48 (1999).
[23] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **116**, 3777 (2002).
[24] For the smaller clusters the clustering coefficients are not that much larger than random simply because these small networks are so highly connected.
[25] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[26] A. L. Barabási and E. Ravasz, Physica (Amsterdam) **299A**, 559 (2001).
[27] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, cond-mat/0112143.
[28] M. E. J. Newman, Phys. Rev. E **64**, 025102 (2001).
[29] H. Jeong, Z. Neda, and A. L. Barabási, cond-mat/0104131.
[30] This dependence of the degree on the energy of a minimum combines with an energetic distribution of minima, which is roughly Gaussian at the most common energies [41,42] but with a low-energy tail that falls off less rapidly, to give the power law tail of the degree distribution.
[31] J. P. K. Doye, D. J. Wales, and M. A. Miller, J. Chem. Phys. **109**, 8143 (1998).
[32] A. Scala, L. A. Nunes Amaral, and M. Barthélémy, Europhys. Lett. **55**, 594 (2001).
[33] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[34] F. Calvo, J. P. K. Doye, and D. J. Wales, J. Chem. Phys. **116**, 2642 (2002).
[35] J. P. K. Doye (unpublished).
[36] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys. Rev. E **64**, 046135 (2001).
[37] B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong, Phys. Rev. E **65**, 027103 (2002).
[38] E. I. Shakhnovich and A. M. Gutin, Europhys. Lett. **9**, 569 (1989).
[39] D. Kohen and F. H. Stillinger, Phys. Rev. E **61**, 1176 (2000).
[40] G. Adam and J. H. Gibbs, J. Chem. Phys. **43**, 139 (1965).
[41] F. Sciortino, W. Kob, and P. Tartaglia, Phys. Rev. Lett. **83**, 3214 (1999).
[42] S. Büchner and A. Heuer, Phys. Rev. E **60**, 6507 (1999).