Small-world properties emerge in highly compartmentalized networks with intermediate group sizes and numbers

Louie H. Yang* and Matthew D. Holland†
Center for Population Biology, University of California, One Shields Avenue, Davis, California 95616, USA
(Received 20 January 2005; published 1 December 2005)

Many recent studies have focused on two statistical properties observed in diverse real-world networks: the small-world property and compartmentalization [D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998); M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. **99**, 7821 (2002)]. Models that include group affiliations have been shown to produce networks with high clustering coefficients, a necessary condition for small-world properties [M. E. J. Newman, Phys. Rev. E, **68**, 026121 (2003); M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003)]. However, the consequences of varying the number and size of groups in a network are not well understood. In order to investigate the consequences of group organization, we examined sets of networks that varied simultaneously in the size and number of groups, while maintaining the same overall size and average degree. Here we show that the small-world property arises in maximally compartmentalized and clustered networks that occur in the intermediate region between few, very large groups and many, very small groups.

DOI: 10.1103/PhysRevE.72.067101 PACS number(s): 89.75.Hc, 89.75.Fb, 87.23.−n

I. INTRODUCTION

Increasingly, network models are being used to represent and analyze the structure of interactions in real world systems. For example, network models have been employed to describe diverse social networks [1–5], engineering networks [6,7], the internet [8], and various biological systems, including metabolic networks [9], epidemic spread in populations [10–13], metapopulations [14], and ecological interaction webs [15–21]. Network representations provide a common framework for the analysis of structural properties in large complex systems. This structural approach is especially useful in systems where a complete dynamical description is not possible, but an analysis of structural properties may yield insights into some aspects of network dynamics [3].

Intriguingly, diverse real world networks often demonstrate common structural properties, including the "small world property" [7] and compartmentalization [16]. The small world property is defined by (1) short characteristic path lengths (L) similar to random networks with the same number of vertices and edges, where L is defined as the mean of all pairwise shortest paths between each vertex and all other vertices in the network, and (2) high clustering coefficients (γ) much greater than random networks with the same number of vertices and edges, where γ is defined as the probability that two vertices within a network that are connected to a randomly chosen third vertex, are themselves connected. This property provides a mathematical foundation for the so-called "six degrees of separation" observation that short interaction chains exist in relatively sparse networks.

The compartmentalization property describes networks where a group of vertices within a network is highly connected, but connections between different groups are sparse. Compartments have been described as modules, communities, subsystems and groups within various network contexts. For example, compartmentalization has been identified in ecological foodwebs [16,19,20,22], where groups may correspond with habitat boundaries [23,24]. In social networks where individuals have group affiliations, interactions may be more common within groups than between groups [2,16]. Such group affiliations are sufficient to produce highly clustered networks [1,5,25], and this group organization suggests a simple explanation for observations of highly clustered real-world networks. However, the consequences of group size and number on the compartmentalization and small world properties of networks are not well understood.

In order to investigate how the size and number of groups in networks influences compartmentalization and the small world property, we analyzed the structural properties of simulated networks using a simple model of a compartmentalized system. This system is represented by a population of N vertices separated into C equal-sized groups (Fig. 1). The probability of an edge between vertices of the same group is w, and the probability of any edge between vertices of different groups is b, where $w \gg b$. In models where $w \gg b$, network representations appear to be composed of C well-

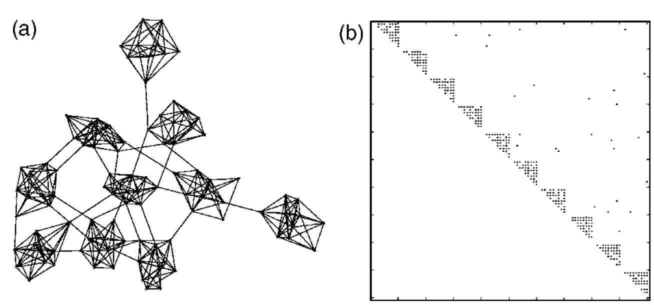

FIG. 1. (a) A graph representation of a highly compartmentalized network of 100 vertices separated into 10 equal-sized groups. (b) The adjacency matrix of the same network.

*Email address: lhyang@ucdavis.edu
†Email address: mdholland@ucdavis.edu

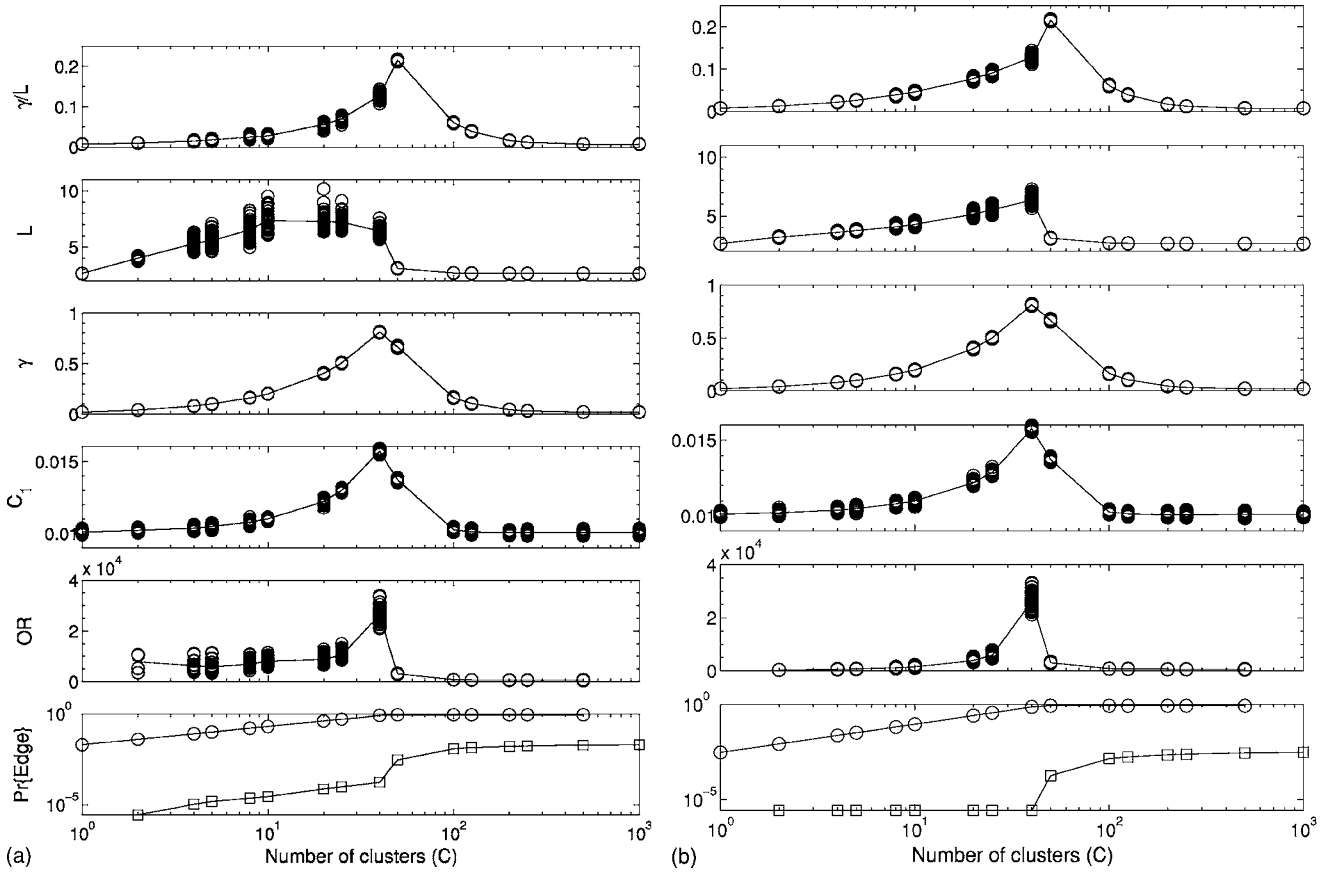

FIG. 2. Structural properties of networks that vary simultaneously in group size and number, while maintaining constant overall size ($N=1000$) and average degree ($k=20$). (a) Where $w < w_{max}$, b was chosen so that approximately half of the graphs were connected. (b) Where $w < w_{max}$, b was fixed at 1.77×10^{-4}. Y-axis labels as follows: Pr{Edge} represents the probabilities of edges, b (between-compartment; squares) and w (within-compartment; circles) used in network simulations; OR is the odds ratio measure of compartmentalization [Eq. (6)]; C_1 [23] [Eq. (7)]; γ is the clustering coefficient; L is the characteristic path length [Eq. (5)]; γ/L is the ratio of the clustering coefficient to characteristic path length.

connected subsystems, connected by a random graph of sparse inter-group edges (Fig. 1). While this model does not attempt to replicate many features of diverse real-world networks, it provides a readily interpretable basis for investigations of group size and number.

II. METHODS

We used a computer program [26] to generate sets of networks that varied simultaneously in the number and size of groups using a range of parameter values in C, b and w, while maintaining the same overall size N and average degree k (Fig. 2). Specifically, undirected networks were generated at every evenly divisible group size between N and 1 using methods similar to those of Girvan and Newman [16]. Twenty networks per parameter set were generated with constant overall size ($N=1000$ vertices) and average degree ($k=20$) for numbers of equal-sized groups, $C\in\{1,2,4,5,8,10,20,25,40,50,100,125,200,250,500,1000\}$. Each network was generated by iterating over all $\binom{1000}{2}$ pairs of vertices, and adding an edge to the network adjacency matrix with probability w for vertices in the same group, or probability b for vertices in different groups. With these parameters, the extreme values of $C=1$ and $C=1000$ produce random graphs with edge probability $p=0.02$. At these extremes, each individual either constitutes its own group, or all individuals in the population belong to the same group; this results in equivalent, uncompartmentalized random networks at both extremes of C. Unconnected networks (networks in which any vertex is unreachable from any other vertex) were omitted from the analysis.

The parameters along the group number axis [Fig. 2(a)] are chosen to yield the greatest compartmentalization possible for a graph with N vertices, C groups, and $E(M)$ expected edges, given a maximum density of within-group edges, w_{max}. The parameters w and b are related as follows:

$$E(M) = \frac{Nw(N-C) + bN^2(C-1)}{2C}. \quad (1)$$

The parameters are chosen in a two-step process. First, an upper bound is calculated for w:

$$w^{+} = \frac{2E(M)C}{N(N-C)}, \quad (2)$$

and w is defined:

$$w = \begin{cases} w^+ & \text{if } w^+ < w_{max}, \\ w_{max} & \text{otherwise}. \end{cases} \quad (3)$$

The between-group edge probability, b, is then calculated by solving Eq. (1) for b:

$$b = \frac{2CE(M) - Nw(N-C)}{N^2(C-1)}. \quad (4)$$

When $w=w^+$ and $b=0$, it is necessary to choose a nonzero value of b, so that the simulated networks will be connected. For each such C and w, we chose a value of b so that approximately half of the networks generated would be connected. These values were obtained by generating large numbers of networks while varying b over a logarithmically spaced set of values. We then recalculated w according to Eq. (1) in order to keep $E(M)$ constant.

Characteristic path length, L, for a network is defined as the arithmetic mean of all pairwise shortest paths between each vertex and all other vertices in the network. We calculate this quantity using Dijkstra's algorithm [27] to obtain all pairwise shortest path lengths l_{ij}, and define

$$L = \frac{1}{\binom{N}{2}} \sum_{i<j} l_{ij}. \quad (5)$$

The average clustering coefficient, γ, is the arithmetic mean of the vertex-wise clustering coefficients, γ_i, where γ_i is defined as the ratio of number of edges between vertices in the neighborhood of vertex i to the number of possible edges in the neighborhood of vertex i. The neighborhood of vertex i is the set of all other vertices sharing an edge with vertex i. Thus, γ_i represents the probability that a pair of vertices sharing connections with a third vertex will themselves be connected. The average clustering coefficient provides a measure of local versus global connections in the graph.

The ratio γ/L provides a useful relative measure of the small-world property. In a set of networks, this ratio is maximized in those networks that have a relatively short characteristic path length and a relatively high clustering coefficient. While the statistical properties of this ratio are uninvestigated, it provides a useful heuristic measure to objectively evaluate small world properties within a defined set of networks.

Another natural measurement of compartmentalization is the increase in the probability of an interaction between individuals within the same compartment, relative to the probability of an interaction between individuals in different compartments. This can be expressed as an odds ratio (OR) of the form

$$\mathrm{OR} = \frac{\Pr\{A|B\}\Pr\{A^C|B^C\}}{\Pr\{A^C|B\}\Pr\{A|B^C\}}, \quad (6)$$

where A=“two vertices share an edge,” B=“two vertices are in the same compartment,” and A^C and B^C represent the complements of events A and B, respectively [22]. We estimated *OR* for simulated networks with varying group size and number. The initial group assignments used to generate the networks were assumed to represent the compartments within these networks.

Finally, we report values for a measure of compartmentalization, C_1, from the food web literature [23]. If we define ν_i as the neighborhood of vertex i, then

$$C_1 = \frac{1}{\binom{N}{2}} \sum_{i<j} c_{ij}, \quad (7)$$

where

$$c_{ij} = \frac{\nu_i \cap \nu_j}{\nu_i \cup \nu_j}. \quad (8)$$

C_1 shows a pattern similar to that of γ, which is not surprising given the similarity of form.

III. RESULTS

Investigations in the vast space of possible topologies rely upon “tunable” axes of variation to delimit sets of comparable networks. Our investigation contributes to previous efforts by considering a novel axis representing variation in the number and size of groups.

Our results show that compartmentalization, clustering and the small world property are maximized in a narrow intermediate region between few, very large groups and many, very small groups (Fig. 2). Two measures of compartmentalization reach an abrupt maximum in networks of intermediate group sizes and numbers. While both clustering coefficients and characteristic path lengths are maximized in networks of high compartmentalization, differences in the behavior of these properties along this axis result in a narrow region of relatively high clustering and relatively short characteristic path lengths. The overall pattern of these results is consistent across a wide range of parameter values in b and w [Fig. 2(b)]. These results are related to Newman's finding that increasing clustering leads to a reduction in the size of the largest connected component for graphs with compartmental structure and constant k [25]. These results suggest that the size and number of groups in networks may influence both compartmentalization and the small world property in model networks.

IV. DISCUSSION

The small world property arises in networks with intermediate group numbers and sizes because these networks balance the exposure of individuals to both global and local interactions. This result suggests that some compartmental network organizations may provide a general basis for the observation of small world properties in many real-world networks. However, not all compartmental network organizations lead to small world properties. For example, some “rewired ring” models begin as highly clustered components which are then connected to each other by randomly removing within-compartment edges and establishing new connections between the disconnected vertices and randomly chosen vertices in the graph with probability ϕ [7]. Larger values of ϕ correspond to an increased incidence of global connections at the expense of local ones. Along an axis of increasing ϕ, compartmentalization decreases monotonically, but the small world property arises at intermediate values of ϕ.

While many prior investigations have focused on the dynamical consequences of small world topologies, studies are increasingly investigating the factors underlying these network structures. For example, Newman and Park [1] present analyses of several real-world social networks in which compartmental network structures influence both the clustering property and degree correlations between adjacent vertices (assortativity). Our findings are consistent with these data in emphasizing the importance of compartmental structures in network properties, and may contribute to previous studies by suggesting a range of groups organizations that are more likely to lead to highly clustered networks, which facilitate small world topologies. Exploring fundamental explanations for the structure of group organizations in real-world networks may provide a mechanistic approach to explain some observations of small worlds in diverse systems, although these explanations are likely to be system-specific. However, this investigation suggests that factors influencing the size and number of groups in real-world networks may have important effects on small world properties.

This analysis also suggests some potential limitations for investigations of real-world networks. While many network properties may be influenced by group structure, the observed group structure is highly dependant on the scale of the network boundaries. The boundaries of real-world networks are often indistinct, and the appropriate scale for investigating the compartmentalization in these networks is unclear. For this reason, group organization may not be apparent at the scale or resolution of many existing datasets. For example, ecological food webs are complex systems that span heterogeneous environments. While many food web studies have focused on communities with relatively distinct habitat boundaries, such as islands or lakes, several recent studies have emphasized the importance of occasional interhabitat subsidies that may link distant food webs [28]. Alternatively, real-world data sets may describe networks with few large groups, although these groups may be part of a larger compartmental system when viewed at a dynamically relevant scale.

ACKNOWLEDGMENTS

We would like to thank A. King, D. Callaway, A. Hastings, C. Hom, V. Rudolf, and R. Yang for helpful suggestions and comments. This work was supported by the National Science Foundation Reseach Training Grant "Nonlinear dynamics in biology" (NSF DBI-9602226).

[1] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).
[2] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).
[3] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[4] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[5] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, Proc. Natl. Acad. Sci. U.S.A. **99**, 2566 (2002).
[6] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[7] D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton University Press, Princeton, NJ, 1999).
[8] L. A. Adamic, in *Proceedings of the ECDL'99 Conference* (Springer-Verlag, New York, 1999).
[9] A. Wagner and D. A. Fell, Proc. R. Soc. London, Ser. B **268**, 1803 (2001).
[10] J. L. Wylie and A. Jolly, Sex Transm. Dis. **28**, 14 (2001).
[11] M. Boots and A. Sasaki, Proc. R. Soc. London, Ser. B **266**, 1933 (1999).
[12] M. J. Keeling, Proc. R. Soc. London, Ser. B **266**, 859 (1999).
[13] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[14] D. Urban and T. Keitt, Ecology **82**, 1205 (2001).
[15] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[16] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[17] J. A. Dunne, R. J. Williams, and N. D. Martinez, Proc. Natl. Acad. Sci. U.S.A. **99**, 12917 (2002).
[18] J. A. Dunne, R. J. Williams, and N. D. Martinez, Ecol. Lett. **5**, 558 (2002).
[19] J. Camacho, R. Guimerà, and L. A. N. Amaral, Phys. Rev. Lett. **88**, 228102 (2002).
[20] J. M. Montoya and R. V. Solé, J. Theor. Biol. **214**, 405 (2002).
[21] R. V. Solé and J. M. Montoya, Proc. R. Soc. London, Ser. B **268**, 2039 (2001).
[22] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, Nature (London) **426**, 282 (2003).
[23] S. L. Pimm and J. H. Lawton, J. Anim. Ecol. **49**, 879 (1980).
[24] J. R. E. Critchlow and S. C. Stearns, Am. Nat. **120**, 478 (1982).
[25] M. E. J. Newman, Phys. Rev. E **68**, 026121 (2003).
[26] Network generation and analysis software was written in C by M.D.H., and can be obtained from the authors on request.
[27] R. Sedgewick, *Algorithms in C, Part 5: Graph Algorithms*, 3rd ed. (Addison-Wesley Professional, Boston, 2001).
[28] G. A. Polis and S. D. Hurd, in *Food Webs: Integration of Patterns and Dynamics*, edited by G. A. Polis and K. O. Winemiller (Chapman and Hall, New York, 1996), pp. 275–285.