International Journal of Modern Physics C
Vol. 19, No. 6 (2008) 939–946

World Scientific
www.worldscientific.com

CLUSTER-DENSE NETWORKS

MOSHE SIPPER[*]
Department of Computer Science, Ben-Gurion University
Beer-Sheva 84105, Israel
sipper@cs.bgu.ac.il

Received 10 September 2007
Accepted 20 January 2008

Small-world networks, exhibiting short nodal distances and high clustering, and scale-free networks, typified by a scale-free, power-law node-degree distribution, have been shown to be widespread both in natural and artificial systems. We propose a new type of network — *cluster-dense network* — characterized by multiple clusters that are highly intra-connected and sparsely inter-connected. Employing two graph-theoretic measures — local density and relative density — we demonstrate that such networks are prevalent in the world of networks.

Keywords: Small-world networks; scale-free networks; social networks; graph measures.

PACS No.: 89.75.-k.

Over the past decade complex networks, both natural and artificial, have moved into the center stage of scientific research. It has been shown that such diverse networks as the World Wide Web, power grids, and the brain have surprising structural commonalities, belonging to one or both of two categories: small-world networks,[1] and scale-free networks.[2]

Small-world networks are highly clustered, much like regular lattices, yet have small characteristic path lengths, like random graphs. This is evidenced by focusing on two graph properties: *average path length* L, and *clustering coefficient* C. L measures the average separation between two vertices in the graph, and is defined as the number of edges in the shortest path between two nodes, averaged over all pairs of nodes. C measures the cliquishness of a typical neighborhood: the average fraction of pairs of neighbors of a node that are also neighbors of each other. More precisely, assume node i in the network has k_i edges connecting it to k_i other nodes. These nodes are all neighbors of node i and at most $k_i(k_i - 1)/2$ edges can exist between them. Denote by E_i the number of these edges that actually exist. The clustering coefficient C_i of node i is then defined as the ratio between E_i and the

[*]URL:www.moshesipper.com.

total number of possible edges, i.e., $C_i = 2E_i/(k_i(k_i - 1))$. The clustering coefficient C of the whole network is the average C_i over all i. While random networks[3] display both low L and C values (i.e., poorly clustered with short nodal distances) and regular lattices exhibit both high L and C values (i.e., highly clustered with long nodal distances), small-world networks are characterized by low L and high C: They are highly clustered yet thanks to "shortcuts" nodal distances are kept low. Many examples of such networks are now known to exist, e.g., movie actors,[1] power grids,[1] nervous systems,[1] human language,[4] electronic circuits,[5] and metabolic networks.[6]

Scale-free networks focus on another graph property, *degree distribution*, defining $P(k)$ as the probability that a vertex in the network has k neighbors (degree k).[2] In a regular lattice all nodes have the same number of edges and so plotting the degree distribution — $P(k)$ vs. k — would result in a single sharp spike (delta distribution). A completely random network obeys a Poisson distribution.[7] Scale-free networks are characterized by a scale-free, power-law degree distribution: $P(k) \sim k^{-\gamma}$. Such networks comprise a multitude of low-degree nodes and relatively few high-degree nodes (usually called "hubs"; Fig. 1). As with their small-world brethren, over the past few years many networks have been shown to be scale-free, e.g., the World Wide Web,[2,8] human sexual contacts,[9] brain functional networks,[10] and the US airline routing map.[7] The latter is an illustrative case in point as it differs drastically from the US roadmap. In the airline routing map nodes represent airports and links represent direct flights between them. The degree distribution is scale-free: a few hubs with flights to most US cities, and a vast majority of nodes with but a few links connecting them to hubs. The US roadmap, on the other hand, is fairly uniform: each major city has at least one link to the highway system but no city is served by hundreds of highways. It is thus an exponential network (Poisson distribution), with connectivity peaking at an average value and then decaying exponentially.[7]

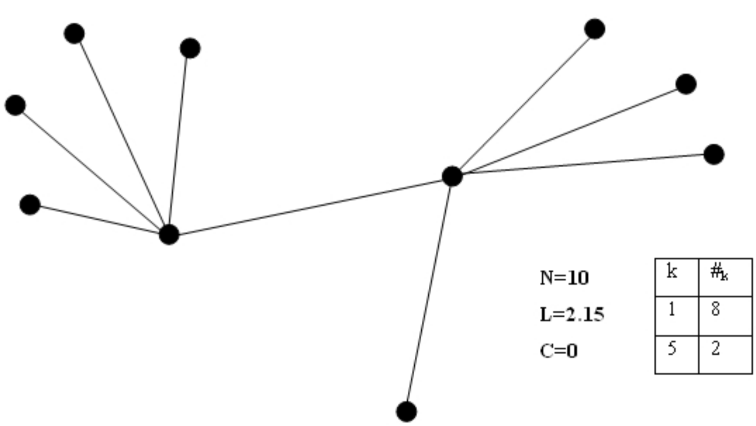
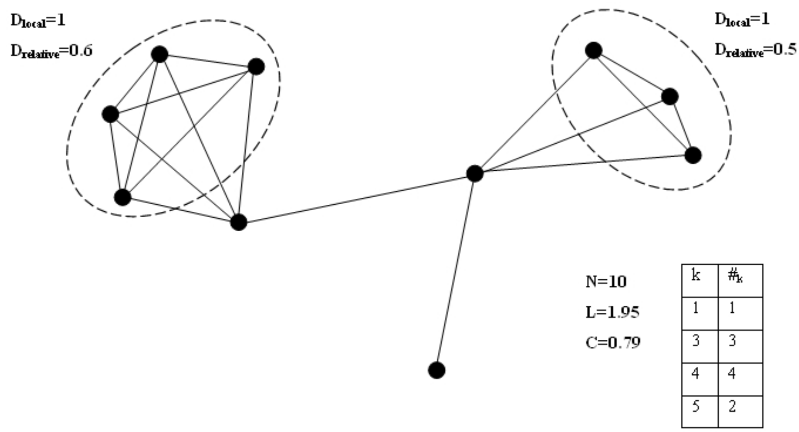
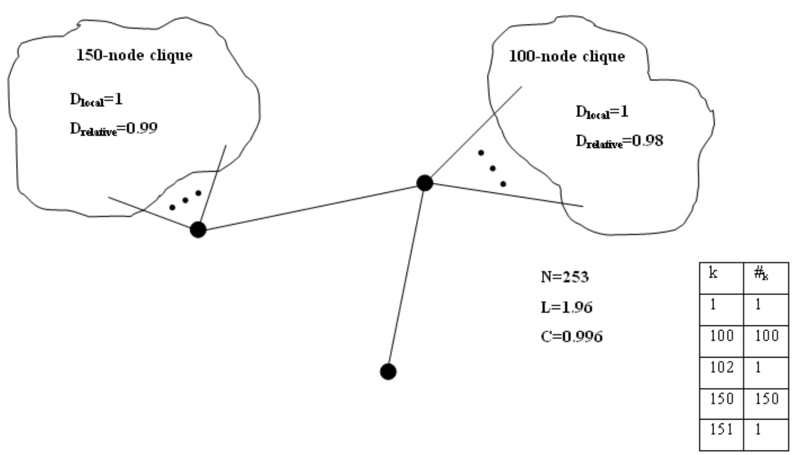


Fig. 1. A scale-free network with $N = 10$ nodes. Shown also are average path length L, clustering coefficient C, and degree distribution (k: node degree, $\#_k$: number of nodes with degree k).



(a)



(b)

Fig. 2. (a) Network derived from Fig. 1 by adding edges such that two clusters are formed. (b) Network derived from (a) by increasing the size of the two clusters, such that the left one becomes a 150-node clique (fully connected subgraph) and the right one becomes a 100-node clique. Note that while the network of Fig. 1 exhibits a typical scale-free degree distribution, networks (a) and (b) do not. Binning similar-degree nodes together, (b) is in fact quite an *atypical* scale-free network, with a positive exponent of approximately 1 rather than the usual negative exponent. D_{local} and D_{relative} are shown for (a) and (b) (the network of Fig. 1 has no clusters).

The network shown in Fig. 1 is typical scale-free: there are eight low-degree nodes, all with degree 1, and two hubs, both with degree 5. Now consider the network of Fig. 2(a), whose nodes are identical to that of Fig. 1, but with added edges that clearly form two clusters. The degree distribution is no longer typical scale-free; this is emphasized in Fig. 2(b). If not scale-free, then perhaps the networks of Fig. 2 are small-world? Indeed, as shown, they exhibit low L and high C —

the two hallmarks of small-world networks. So we seem to have on our hands non-scale-free, small-world networks, which is quite unremarkable: the examples given by Ref. 1 almost a decade ago fall into this exact same category — small-world, but with probability of finding a highly connected node (that is, a large k) decreasing exponentially with k.[2]

My argument herein is that we are, nonetheless, faced with a new category of network, unlike small-world networks studied so far. The networks of Fig. 2 display a characteristic form of clustering that merits special attention. In order to quantify their uniqueness, I employ two graph-theoretic measures: *local density*, D_{local}, a measure of a cluster's inner density of edges, and *relative density*, D_{relative}, a measure of a cluster's "introvertness".[11,12] Given a graph, a cluster c is a connected subgraph. The *internal degree*, deg_{int}, of c is the number of edges that have both endpoints in c; the *external degree*, deg_{ext}, of c is the number of edges that have only one endpoint in c. The local density, D_{local}, of cluster c with k nodes is the ratio of the internal degree of the cluster to the maximum possible — $k(k-1)/2$:

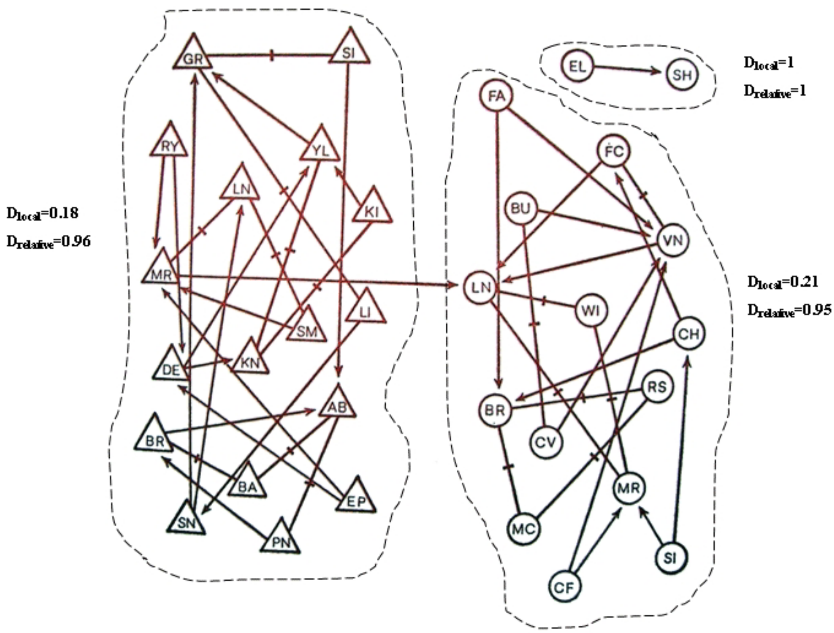
$$D_{\text{local}}(c) = \frac{2deg_{\text{int}}(c)}{k(k-1)}\,.$$

By convention, the density of an empty- or a single-vertex cluster is zero. The relative density, D_{relative}, of cluster c is the ratio of internal edges to the total number of edges impinging upon the cluster:

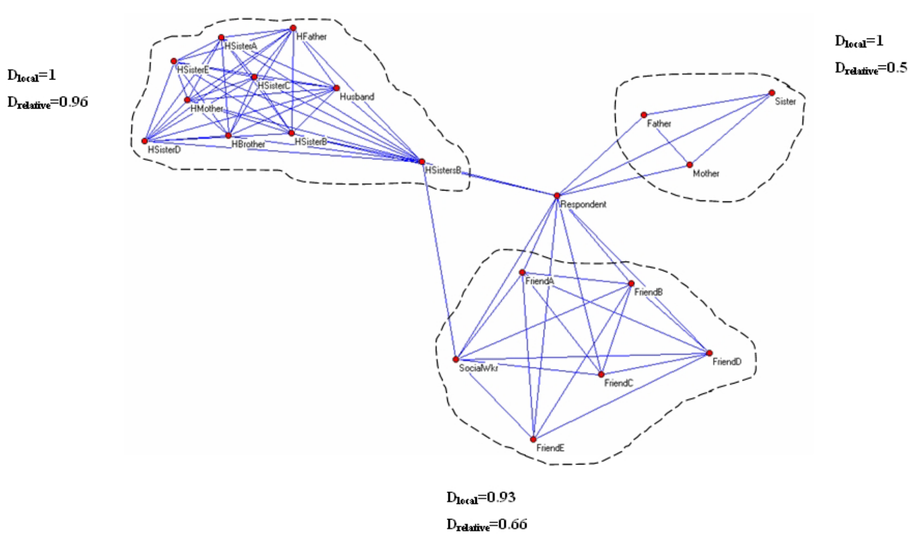
$$D_{\text{relative}}(c) = \frac{deg_{\text{int}}(c)}{deg_{\text{int}}(c) + deg_{\text{ext}}(c)}\,.$$

Figure 2 shows these measures for the networks in question (the scale-free network of Fig. 1 has no clusters). As can be seen both D_{local} and D_{relative} are high ($D_{\text{local}} \gg 0$, $D_{\text{relative}} \gg 0$), exemplifying what we call *cluster-dense networks*: networks rich in relatively dense clusters, which are also "introverted," i.e., have few connections outside the cluster.

Note that the classes of cluster-dense and small-world networks, while overlapping, are not identical. A small-world network need not necessarily embody any natural clustering: A high clustering coefficient C, evidencing a node's local neighborhood clustering, does not in itself entail the global existence of clusters (two people may share many friends, who in turn share many friends, etc — without there being naturally occurring clusters). Conversely, a cluster-dense network is not necessarily small-world: Watts[13] (see also Ref. 14) carries out an in-depth analysis of the small-world phenomenon, along the way asking what graph has the lowest possible characteristic path length and what is the most clustered graph possible. The former is attained by a random graph while toward the latter Watts proposes the "connected caveman graph": $n/(k+1)$ isolated cliques ("caves") of size $k+1$ nodes, where one edge from each clique is extracted and connected to a neighboring clique such that all cliques form a connected loop. This network has $C \approx 1$ and large L (i.e., L increases linearly with network size n), the latter characteristic rendering it non-small-world. The network is, of course, cluster-dense.

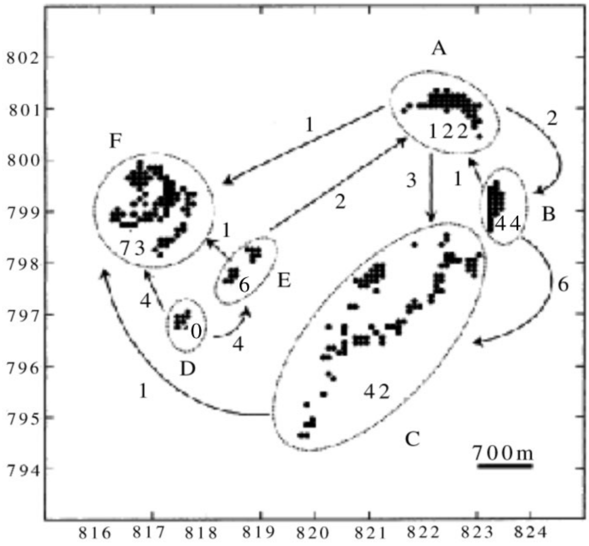


(a)

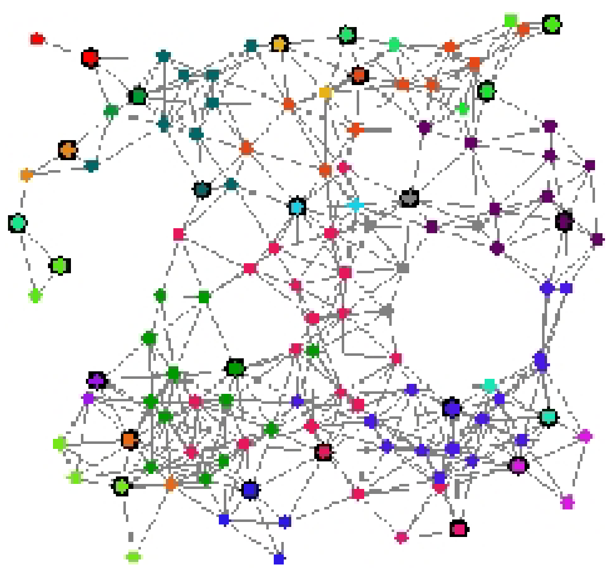


(b)

Fig. 3. (a) Friendship choices among fourth graders.[15] Triangles represent boys and circles represent girls. Three clusters are clearly visible. (We ignore the graph's being directional, as we are interested in any social interaction, regardless of its directionality.) (b) Social ties surrounding a homeless woman (*Respondent*) (graph appears in Ref. 16 [reprinted with permission], based on data reported by Ref. 17). Three clusters are distinctly visible.



(a)



(b)

Fig. 4. (a) Distribution of adults of *Zygaena filipendulae* (a moth species) and observed movements between populations [reprinted with permission].[18] Interestingly, here the inter-cluster edges represent nodal movements (migration). Computing D_{local} and D_{relative} can be done by ascribing the (relatively few) transferred nodes to their cluster of origin. Six clusters are present, each with $D_{\text{local}} = 1$ (all moths may interact within a cluster) and $D_{\text{relative}} \approx 1$ (number of migrating moths $\ll$ number of moths within cluster). (b) Ad-hoc wireless network, artificially constructed to maximize the product of D_{local} and D_{relative} of clusters in a dynamically forming network [reprinted with permission].[11]

The main question is how prevalent are cluster-dense networks: Are they worthy of study as a class unto its own? Figures 3 and 4 provide some evidence in favor of an affirmative reply. The first two examples (Fig. 3) display typical cluster-dense networks that arise naturally amongst humans: a friendship network and a support network. This demonstrates one method by which such networks are formed: natural aggregation, as occurs with humans and, more generally, in the faunal world. Figure 4(a) shows a faunal cluster-dense network: moth populations. Figure 4(b) shows a man-made cluster-dense network — an ad-hoc wireless network that was designed through algorithmic aggregation: by seeking to maximize the product of D_{local} and D_{relative} of clusters in a dynamically forming network.[11]

The literature contains reports of other cluster-dense networks:

- The social human milieu probably provides ample examples, due to the human instinct to socialize; e.g., Freeman[16] discusses cluster-dense networks of visiting patterns among households in El Cerrito, New Mexico in 1940, and family clusters in Atirro, Costa Rica.
- Turning to artificial networks, the World Wide Web exhibits communities (of pages and links) that are typical cluster-dense.[19,20]
- Animals tend to aggregate into flocks, schools, herds, packs, bevies, swarms, hordes, troops, and so forth, many of which also display characteristics of cluster-dense networks (e.g., Fig. 4(a)). As (another) example, consider penguins in the Galápagos islands,[21] which form an interaction network that is also cluster-dense: high local density per island (high D_{local}), and high relative density between islands (high D_{relative}) due to a physical barrier (the Pacific Ocean). Another example is that of butterfly populations.[22,23]

Average path length, clustering coefficient, and degree distribution, often being easy to compute, renders small-world and scale-free networks easily identifiable.[a] Cluster-dense networks are harder to detect since identifying clusters is generally intractable, the basic questions related to graph partitioning and thus being NP-complete.[20,24] Nonetheless, many clustering algorithms can provide good approximate solutions; moreover, humans, quite proficient at the clustering task, may provide a quick resolution (albeit for relatively small-sized networks, networks that afford good visualization, or both).

Cluster-dense networks probably arise often both in natural and man-made systems. Their dynamics may be quite different than those of other proposed network models and could thus benefit from special consideration.

Acknowledgment

I thank Marco Tomassini and the anonymous referee for their helpful remarks.

[a]This is not always true. As noted by Watts,[13] for large social networks even local parameters like k and C can be hard to estimate.

References

1. D. J. Watts and S. H. Stogatz, *Nature* **393**, 440 (1998).
2. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
3. P. Erdös and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960); B. Bollobás, *Random Graphs*, 2nd edn. (Cambridge University Press, Cambridge, 2001).
4. R. Ferrer i Cancho and R. V. Solé, *Proc. R. Soc. Lond. B* **268**, 2261 (2001).
5. R. Ferrer i Cancho, C. Janssen and R. V. Solé, *Phys. Rev. E* **64**, 046119 (2001).
6. A. Wagner and D. A. Fell, *Proc. R. Soc. Lond. B* **268**, 1803 (2001).
7. X. F. Wang and G. Chen, *IEEE Circuits and Systems Magazine* **3**, 6 (2003).
8. A.-L. Barabási, R. Albert and H. Jeong, *Physica A* **281**, 69 (2000).
9. F. Liljeros, C. R. Edling, L. A. Nunes Amaral, H. E. Stanley and Y. Aberg, *Nature* **411**, 907 (2001).
10. V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki and A. V. Apkarian, *Phys. Rev. Lett.* **94**, 018102 (2005).
11. S. E. Schaeffer, S. Marinoni, M. Särelä and P. Nikander, in *Proc. 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networkds*, Reston, VA, USA, Vol. 2 (2006), pp. 667–672.
12. M. Mihail, C. Gkantsidis, A. Saberi and E. Zegura, On semantics of internet topologies, Technical Report GIT-CC-02-07, Georgia Institute of Technology (2002).
13. D. J. Watts, *Am. J. Sociol.* **105**, 493 (1999).
14. S. Virtanen, Properties of nonuniform random graph models, Technical Report A77, Laboratory for Theoretical Computer Science, Helsinki University of Technology (2003).
15. J. L. Moreno, *Who Shall Survive?* (Nervous and Mental Disease Publishing Company, Washington DC, 1934).
16. L. C. Freeman, *J. Soc. Struct.* **1** (2000).
17. J. C. Mitchell, *Connections* **17**, 16 (1994).
18. R. Menéndez, D. Gutiérrez and C. D. Thomas, *Ecol. Entomol.* **27**, 317 (2002).
19. J. Kleinberg and S. Lawrence, *Science* **294**, 1849 (2001).
20. G. W. Flake, S. Lawrence, C. L. Giles and F. M. Coetzee, *IEEE Comput.* **35**, 66 (2002).
21. H. Vargas, C. Lougheed and H. Smell, *Ibis* **147**, 367 (2005).
22. D. Gutiérrez, C. D. Thomas and J. L. León-Cortés, *Oecologia* **121**, 506 (1999).
23. C. Schneider, J. Dover and G. L. A. Fry, *Ecol. Entomol.* **28**, 219 (2003).
24. J. Šíma and S. E. Schaeffer, in *Proc. 32nd International Conference on Current Trends in Theory and Practice of Computer Science*, Vol. 3831, Lecture Notes in Computer Science (Springer Verlag, Berlin-Heidelberg, 2006), pp. 530–537.