What led to the pervasiveness of hybrids between *Cx. pipiens* and *Cx. molestus* in North America, but not in Europe and Africa, still remains to be determined. In southernmost Europe, we identified two populations with a few hybrid individuals, as well as populations with pure *Cx. pipiens* signatures and populations with a mix of pure *Cx. pipiens* and pure *Cx. molestus* signatures (Fig. 2). Indeed, previous allozyme- based studies indicated the existence of populations in Italy with a mix of the two forms (*26*) but a very low rate of hybridization (1%), probably because of their different mating behaviors (*26*). The rarity of southern European hybrids and our failure to find hybrids in northern Europe may be due to their low fitness and inability to diapause. Importantly, the introduction to the United States of separate populations of *Cx. pipiens* and *Cx. molestus* that later hybridized, or of hybrids from southern Europe, has led to abundant and ubiquitous hybrid forms that survive the rigors of northern winters.

It is now clear that models derived from the U.S. epidemic of WNV (*28*) may not be applicable to Eurasia, and vice versa (*29*). A major factor in all recent outbreaks (Romania 1996, Russia 1999, and United States 1999) is the involvement of mosquitoes in the *Cx. pipiens* complex as the primary vectors (*8*, *30*). Unlike European *Cx. pipiens*, U.S. *Cx. pipiens* appears to bite readily both avian hosts and humans (*2*, *31*). Here we have shown that, across the northeastern United States, a large proportion of individuals are hybrids of human-biter and bird-biter forms. In combination with susceptible migrating birds and highly concentrated human populations in U.S. cities and suburbs, the prevalence of such bridge vectors that readily transmit the virus among and between avian hosts and humans could have created the current epidemic conditions.

The present study suggests that changes in vectorial capacity and the creation of new efficient vectors may occur with new introductions. In particular, the arrival of hybrid American forms in northern Europe has the potential to radically change the dynamics of WNV in Europe.

References and Notes
1. E. B. Vinagradova, *Culex pipiens pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance, and Control* (Pensoft, Moscow, 2000).
2. A. Spielman, *Ann. N.Y. Acad. Sci.* **951**, 220 (2001).
3. M. J. Turell, M. L. O'Guinn, D. J. Dohm, J. W. Jones, *J. Med. Entomol.* **38**, 130 (2001).
4. K. A. Bernard *et al.*, *Emerg. Infect. Dis.* **7**, 679 (2001).
5. V. L. Kulasekera *et al.*, *Emerg. Infect. Dis.* **7**, 722 (2001).
6. D. J. Dohm, M. R. Sardelis, M. J. Turell, *J. Med. Entomol.* **39**, 640 (2002).
7. R. S. Nasci *et al.*, *Emerg. Infect. Dis.* **7**, 742 (2001).
8. C. G. Hayes, *Ann. N.Y. Acad. Sci.* **951**, 25 (2001).
9. R. E. Harbach, C. Dahl, G. B. White, *Proc. Entomol. Soc. Wash.* **87**, 1 (1985).
10. T. Guillemaud, N. Pasteur, F. Rousset, *Proc. R. Soc. London Ser. B.* **264**, 245 (1997).
11. K. L. Knight, *Supplement to the Catalog of the Mosquitoes of the World (Diptera: Culicidae)*, T. S. Foundation, Ed. (Entomological Society of America, College Park, MD, 1978), supplement to vol. 6.
12. K. Byrne, R. A. Nichols, *Heredity* **82**, 7 (1999).
13. R. E. Harbach, B. A. Harrison, A. M. Gad, *Proc. Entomol. Soc. Wash.* **86**, 521 (1984).
14. P. F. Mattingly *et al.*, *Trans. R. Entomol. Soc. Lond.* **102**, 331 (1951).
15. D. X. Zhang, G. M. Hewitt, *Mol. Ecol.* **12**, 563 (2003).
16. N. Keyghobadi, M. A. Matrone, G. D. Ebel, L. D. Kramer, D. M. Fonseca, *Mol. Ecol. Notes* **4**, 20 (2004).
17. D. M. Fonseca, C. T. Atkinson, R. C. Fleischer, *Mol. Ecol.* **7**, 1617 (1998).
18. F. H. Drummond, *Trans. R. Entomol. Soc. Lond.* **102**, 369 (1951).
19. K. Tanaka, K. Mizusawa, E. S. Saugstad, *Contrib. Am. Entomol. Inst.* **16**, 1 (1979).
20. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
21. A. R. Barr, *Am. J. Trop. Med. Hyg.* **6**, 153 (1957).
22. A. J. Cornel *et al.*, *J. Med. Entomol.* **40**, 36 (2003).
23. S. Urbanelli, F. Silvestrini, W. K. Reisen, E. De Vito, L. Bullini, *J. Med. Entomol.* **34**, 116 (1997).
24. L. L. Cavalli-Sforza, F. Cavalli-Sforza, *The Great Human Diasporas: The History of Diversity and Evolution* (Addison-Wesley, Reading, MA, 1995).
25. J. de Zulueta, *Parassitologia* **36**, 7 (1994).
26. S. Urbanelli *et al.*, in *Ecologia, Atti I Congr. Naz. Soc. Ital. Ecol.*, A. Moroni, O. Ravera, A. Anelli, Eds. (Zara, Parma, Italy, 1981), pp. 305–316.
27. C. Chevillon, R. Eritja, N. Pasteur, M. Raymond, *Genet. Res.* **66**, 147 (1995).
28. P. D. Crook, N. S. Crowcroft, D. W. Brown, *Commun. Dis. Public Health* **5**, 138 (2002).
29. Z. Hubalek, *Viral Immunol.* **13**, 415 (2000).
30. H. M. Savage *et al.*, *Am. J. Trop. Med. Hyg.* **61**, 600 (1999).
31. A. G. Richards, *Entomol. News* **52**, 211 (1941).
32. We thank N. Becker, C. Curtis, M. Carroll, G. Ebel, C. Evans, M. Santa Ana Gouveia, L. Kramer, G. O'Meara, F. Noguera, and C. Williams for providing invaluable mosquito samples; J. Smith, M. Matrone, T. Ganguly, and the DNA Sequencing Facility, University of Pennsylvania, for technical assistance; and A. Bhandoola and four anonymous reviewers for comments and valuable suggestions on an earlier version of this manuscript. Supported by a National Research Council Associateship through the Walter Reed Army Institute of Research (D.M.F.), by NIH grant nos. U50/CCU220532 and 1R01GM063258, and by NSF grant no. DEB-0083944. This material reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense.

Superfamilies of Evolved and Designed Networks

Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, Uri Alon*

Complex biological, technological, and sociological networks can be of very different sizes and connectivities, making it difficult to compare their structures. Here we present an approach to systematically study similarity in the local structure of networks, based on the significance profile (SP) of small subgraphs in the network compared to randomized networks. We find several superfamilies of previously unrelated networks with very similar SPs. One superfamily, including transcription networks of microorganisms, represents "rate-limited" information-processing networks strongly constrained by the response time of their components. A distinct superfamily includes protein signaling, developmental genetic networks, and neuronal wiring. Additional superfamilies include power grids, protein-structure networks and geometric networks, World Wide Web links and social networks, and word-adjacency networks from different languages.

Many networks in nature share global properties (*1*, *2*). Their degree sequences (the number of edges per node) often follow a long-tailed distribution, in which some nodes are much more connected than the average

Departments of Molecular Cell Biology, Physics of Complex Systems, and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel.

*To whom correspondence should be addressed at Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: urialon@weizmann.ac.il

(*3*). In addition, natural networks often show the small-world property of short paths between nodes and highly clustered connections (*1*, *2*, *4*). Despite these global similarities, networks from different fields can have very different local structure (*5*). It was recently found that networks display certain patterns, termed "network motifs," at much higher frequency than expected in randomized networks (*6*, *7*). In biological networks, these motifs were suggested to be recurring circuit elements that carry out key information-processing tasks (*6*, *8–10*).

To understand the design principles of complex networks, it is important to compare the local structure of networks from different fields. The main difficulty is that these networks can be of vastly different sizes [for example, World Wide Web (WWW) hyperlink networks with millions of nodes and social networks with tens of nodes] and degree sequences. Here, we present an approach for comparing network local structure, based on the significance profile (SP). To calculate the SP of a network, the network is compared to an ensemble of randomized networks with the same degree sequence. The comparison to randomized networks compensates for effects due to network size and degree sequence. For each subgraph i, the statistical significance is described by the Z score (11):

$$Z_i = (N\mathrm{real}_i - <N\mathrm{rand}_i>)/\mathrm{std}(N\mathrm{rand}_i)$$

where $N\mathrm{real}_i$ is the number of times the subgraph appears in the network, and $<N\mathrm{rand}_i>$ and $\mathrm{std}(N\mathrm{rand}_i)$ are the mean and standard deviation of its appearances in the randomized network ensemble. The SP is the vector of Z scores normalized to length 1:

$$SP_i = Z_i/(\Sigma Z_i^2)^{1/2}$$

The normalization emphasizes the relative significance of subgraphs, rather than the absolute significance. This is important for comparison of networks of different sizes, because motifs (subgraphs that occur much more often than expected at random) in large networks tend to display higher Z scores than motifs in small networks (7).

We present in Fig. 1 the SP of the 13 possible directed connected triads (triad significance profile, TSP) for networks from different fields (12). The TSP of these networks is almost always insensitive to removal of 30% of the edges or to addition of 50% new edges at random, demonstrating that it is robust to missing data or random data errors (SOM Text). Several superfamilies of networks with similar TSPs emerge from this analysis. One superfamily includes sensory transcription networks that control gene expression in bacteria and yeast in response to external stimuli. In these transcription networks, the nodes represent genes or operons and the edges represent direct transcriptional regulation (6, 13–15). Networks from three microorganisms, the bacteria *Escherichia coli* (6) and *Bacillus subtilis* (14) and the yeast *Saccharomyces cerevisiae* (7, 15), were analyzed. The networks have very similar TSPs (correlation coefficient $c > 0.99$). They show one strong motif, triad 7, termed "feed-forward loop." The feedforward loop has been theoretically and experimentally shown
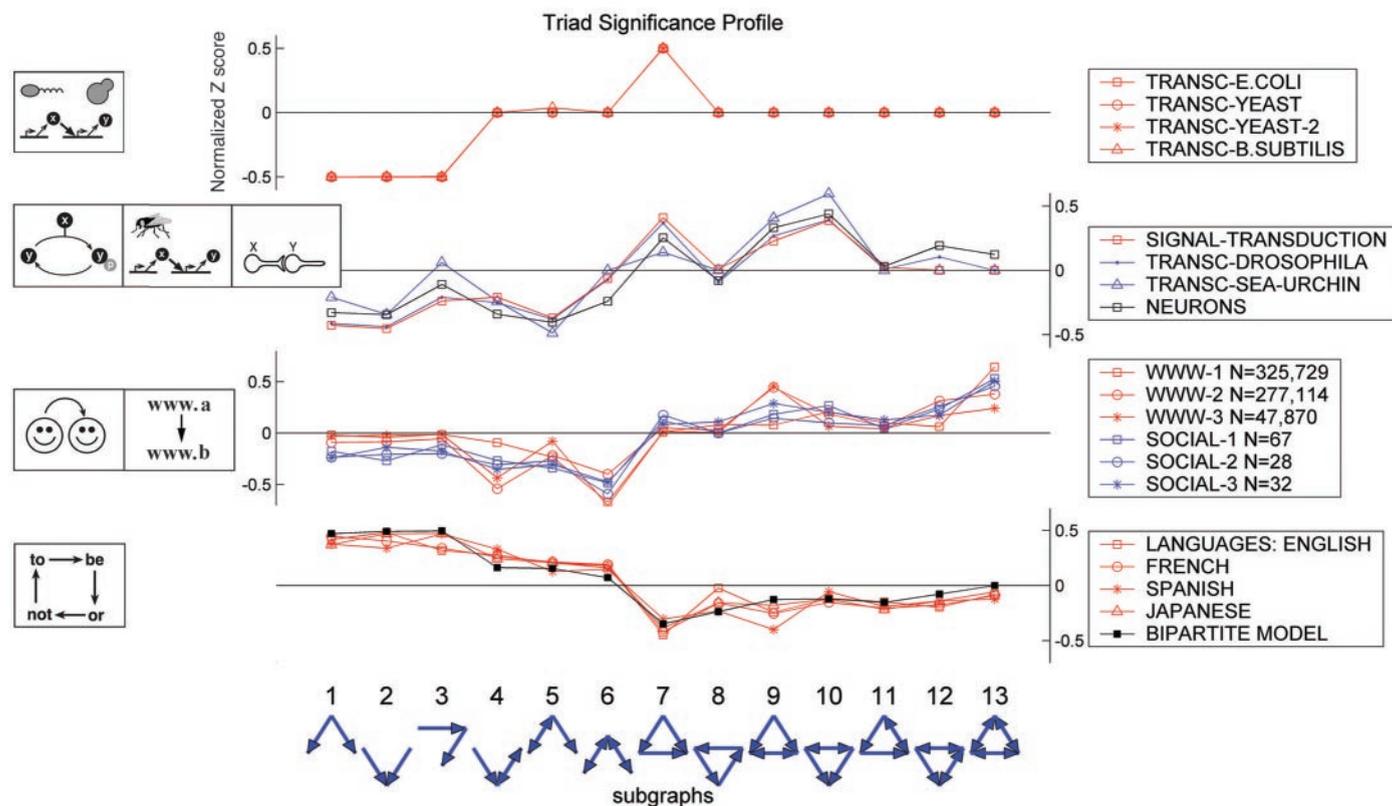
Fig. 1. The triad significance profile (TSP) of networks from various disciplines. The TSP shows the normalized significance level (Z score) for each of the 13 triads. Networks with similar characteristic profiles are grouped into superfamilies. The lines connecting the significance values serve as guides to the eye. The networks are as follows (where N and E are the number of nodes and edges, respectively) (12): (i) Direct transcription interactions in the bacteria *E. coli* (6) (TRANSC-E.COLI $N = 424$, $E = 519$) and *B. subtilis* (14) (TRANSC-B.SUBTILIS $N = 516$, $E = 577$) and in the yeast *S. cerevisiae* [TRANC-YEAST $N = 685$, $E = 1052$ (7) and TRANSC-YEAST-2 $N = 2341$, $E=3969$ (15)]. (ii) Signal-transduction interactions in mammalian cells based on the signal transduction knowledge environment (STKE, http://stke.sciencemag.org/) (SIGNAL-TRANS-DUCTION $N = 491$, $E = 989$), transcription networks that guide development in fruit fly (from the GeNet literature database, www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm) (TRANSC-DROSOPHILA $N = 110$, $E = 307$), endomesoderm development in sea urchin (20) (TRANSC-SEA-URCHIN $N = 45$, $E = 83$), and synaptic connections between neurons in *C. elegans* (NEURONS $N = 280$, $E = 2170$). (iii) WWW hyperlinks between Web pages in the www.nd.edu site (3) (WWW-1 $N = 325729$, $E = 1469678$), pages related to literary studies of Shakespeare (21) (WWW-2 $N = 277114$, $E = 927400$), and pages related to tango, specifically the music of Piazzolla (21) (WWW-3 $N = 47870$, $E = 235441$); and social networks, including inmates in prison (SOCIAL-1 $N = 67$, $E = 182$), sociology freshmen (22) (SOCIAL-2 $N = 28$, $E = 110$), and college students in a course about leadership (SOCIAL-3 $N = 32$, $E = 96$). (iv) Word-adjacency networks of a text in English (ENGLISH $N = 7724$, $E = 46281$), French (FRENCH $N = 9424$, $E = 24295$), Spanish (SPANISH $N = 12642$, $E = 45129$), and Japanese (JAPANESE $N = 3177$, $E = 8300$) and a bipartite model with two groups of nodes of sizes $N_1 = 1000$ and $N_2 = 10$ with probability of a directed or mutual edge between nodes of different groups being $p = 0.06$ and $q = 0.003$, respectively, and no edges between nodes within the same group (BIPARTITE $N = 1010$, $E = 1261$).

to perform signal-processing tasks such as persistence detection (*6*, *10*), pulse generation, and acceleration of transcription responses (*9*). Triad 3, the 3-chain, is an anti-motif (a significantly underrepresented subgraph) corresponding to the shallow architecture of these networks, which have few long cascades (*16*). These networks are "sensory networks," which need to respond within minutes to transient signals such as stresses and nutrients. The minimal time required for response (for the first proteins to be expressed) is indeed on the order of minutes. If the information needs to pass additional steps (a regulator protein needs to be expressed and cross its activation threshold to turn on a gene), then the response time is much longer. The response time in each cascade step has been experimentally and theoretically shown to be on the order of the gene-product lifetime (*8*), often tens of minutes. Thus, these networks are "rate-limited networks," where the desired response times are often as short as the response times of the network components.

In the rate-limited network superfamily, long cascades and feedback loops are rare (*16*). Feedbacks are usually closed by protein-protein interactions and not by transcription (*17*). Purely transcriptional feedback loops may be rare because they are unstable and noisy due to their delays (*18*) or because they can be locked into an irreversible state (*19*), both undesirable properties for sensory transcription networks.

We find a distinct superfamily that includes three types of biological networks (*12*): signal-transduction interactions in mammalian cells based on the Signal Transduction Knowledge Environment (STKE) (*12*), developmental transcription networks that guide development in fruit fly (*12*) and sea-urchin (*20*), and synaptic wiring between neurons in *Caenorhabditis elegans* (*12*). These networks show triads 7, 9, and 10 with positive TSPs, and triads 1, 2, 4, and 5 with negative TSPs (Fig. 1). In contrast to the sensory transcription networks of microorganisms, these networks display two-node feedbacks that regulate or are regulated by a third node (triads 10, 9) and are less biased against cascades (triad 3). The common feature to this superfamily of information-processing networks is that the response time of each step in the network is usually much shorter than the response time required for the biological function of the network. Protein signal-transduction networks often need to respond within an hour or longer, but each interaction can take minutes or less. Cascade steps in developmental networks can have response times of tens of minutes, but the processes they control are much slower, on the order of animal (*19*) cell-division times that can take several hours. Neuronal net-
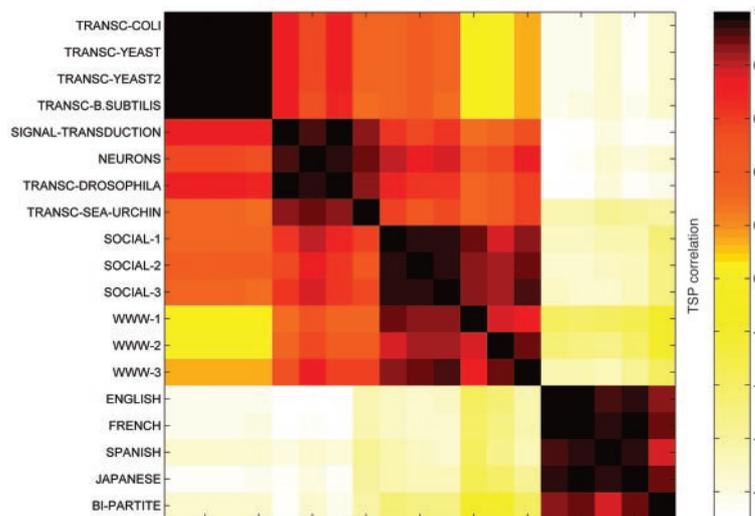
works in *C. elegans* typically need to respond within a second, but neuron response times are shorter than 100 ms. Thus, it appears that this superfamily characterizes biological information-processing networks that are not rate-limited.

Next, we analyzed three WWW networks of hyperlinks between Web pages related to university, literature, or music (*3*, *21*). The TSPs were quite similar ($c = 0.7$ to 0.9). Triads 9, 10, 12, and 13 had the highest TSP values, and 4, 5, and 6 the lowest. The overrepresented triads have many transitive triplet interactions, where if $x{\rightarrow}y$ and $y{\rightarrow}z$ then $x{\rightarrow}z$ (table S2). For example, the overrepresented triad 13, termed "clique," has six transitive interactions, the highest transitivity possible in a triad. The less represented triads such as 6, 8, and 11 are highly intransitive.

We also analyzed three social networks, in which nodes represent people in a group, and edges represent positive sentiment directed from one group member to another, based on questionnaires (*12*, *22*). The TSPs of the three social networks were very similar ($c = 0.92$ to 0.96). Notably, their TSPs were quite close to the TSP of the WWW nets ($c = 0.7$ to 0.9). This suggests that WWW networks and social networks may be part of a superfamily. The tendency of social networks to display transitive interactions and transitive triads is well established (*23*). The similarity of WWW and social networks suggests that classical models of social structural organization (*24*) may also be used to understand WWW structure.

Texts can also be represented as networks (*25*). We studied word-adjacency networks in which each node represents a word and a directed connection occurs when one word directly follows the other in the text. The TSPs of texts from different languages and of
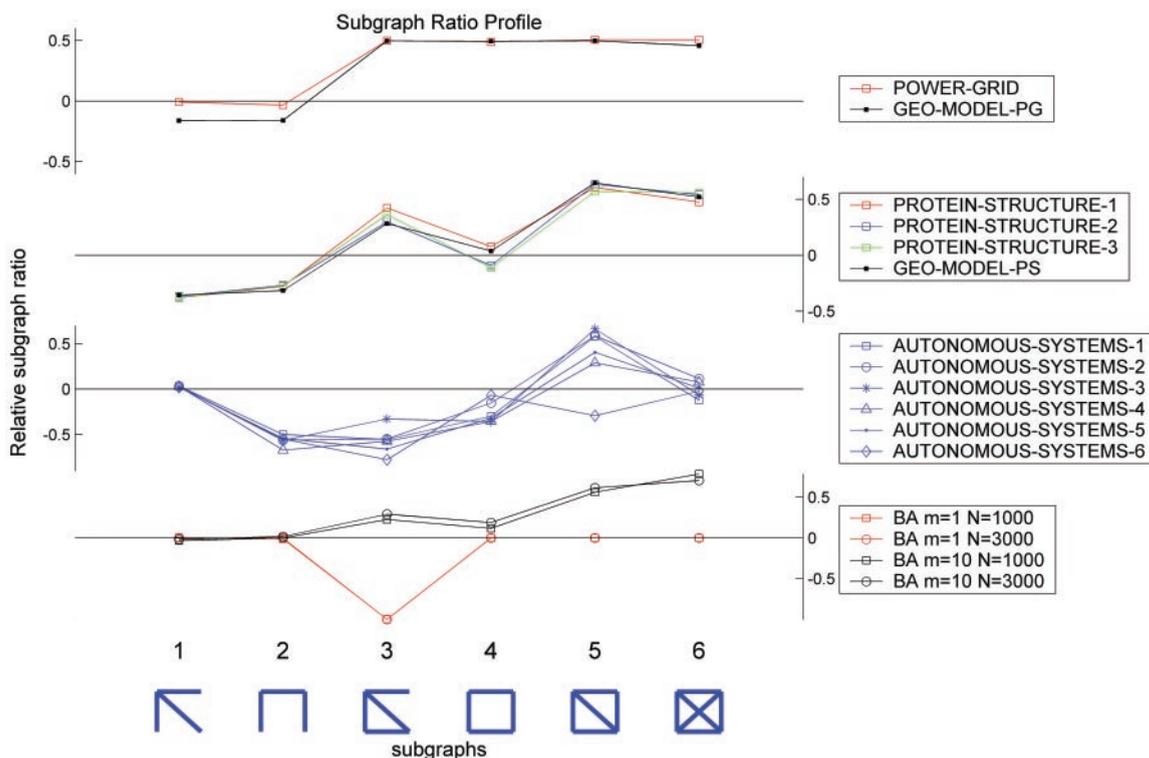
different sizes were similar (Fig. 1 compares texts in English, French, Spanish, and Japanese). The main feature was the underrepresentation of triangle-shaped triads 7 to 13. This is related to the structure of languages in which words belong to categories and a word from one category tends to be followed by one from a different category (*26*). For example, among the most connected words are prepositions, which are usually followed by nouns or articles. Figure 1 also shows the TSP for a model bipartite graph (*12*) in which nodes belong to two groups and connections are formed between these groups and not within the groups. The bipartite model graph shows a TSP that resembles that of the word-adjacency networks.

The similarities between networks can also be visualized by looking at the correlation between the TSPs of different networks (Fig. 2). The correlations can be used to cluster the networks into distinct superfamilies (*27*).

The TSPs display certain conservation relations between subgraph types. For example, networks with excess triangle-shaped triads tend to have a deficit of V-shaped triads. We find that there are several triad conservation rules in networks that conserve the degree sequences of single and mutual edges (SOM Text). As a result, the 13 values of the TSP are determined by only seven degrees of freedom. One can intuitively interpret these conservation laws in terms of reactions that convert V-shaped subgraphs into triangles, preserving the degrees of all participating nodes. Analysis of the reactions occurring in each network allows a compact description of the difference between networks and their randomized counterparts (SOM Text).

We now consider nondirected networks, in which edges have no specified direction-

Fig. 2. The correlation coefficient matrix of the triad significance profiles for the directed networks in Fig. 1.

Fig. 3. The subgraph ratio profile (SRP) for various nondirected networks. The networks are as follows (*12*): (i) The electrical power grid of the western United States (*4*) (POWERGRID $N = 4941$, $E = 6594$) and a geometric model with similar clustering coefficient (GEO-MODEL-PG $N = 5000$, $E = 7499$). (ii) Networks of secondary-structure elements adjacency for several large proteins [structure based on the PDB database (www.rcsb.org/pdb/); the proteins (and their PDB ID) were 1A4J, an immunoglobulin (PROTEIN-STRUCTURE-1 $N = 95$, $E = 213$); 1EAW, a serine protease inhibitor (PROTEIN-STRUCTURE-2 $N = 53$, $E = 123$); and 1AOR, an oxidoreductase (PROTEIN-STRUCTURE-3 $N = 99$, $E = 212$)] and a geometric model with similar clustering coefficient (GEO-MODEL-PS $N = 53$, $E = 136$). (iii) The Internet at the autonomous system level (www.cosin.org) (AUTONOMOUS-SYSTEMS 1 to 6; $N = 3015$, 3522, 4517, 5357, 7956, 10515; $E = 5156$, 6324, 8376, 10328, 15943, 21455). (iv) Networks grown according to the preferential attachment BA model (*3*) with $m = 1$ or $m = 10$ edges per new node (BA $m = 1, 10$; $N = 1000, 3000, 1000, 3000$; $E = 1000, 3000, 9901, 29901$).

ality. Because nondirected networks have only two types of triads (V and triangle), we analyzed the profile of the six types of nondirected connected tetrads (four-node subgraphs). Unlike triads, the normalized Z scores of tetrads show a significant dependence on the network size. Therefore, instead of an SP based on Z scores, we use the abundance of each subgraph i relative to random networks:

$$\Delta_i = \frac{N\mathrm{real}_i - <N\mathrm{rand}_i>}{N\mathrm{real}_i + <N\mathrm{rand}_i> + \varepsilon}$$

where ε ensures that $|\Delta|$ is not misleadingly large when the subgraph appears very few times in both the real and random networks (here $\varepsilon = 4$). The subgraph ratio profile (SRP) is the vector of Δ_i normalized to length 1:

$$SRP_i = \Delta_i / (\Sigma \Delta_i^2)^{1/2}$$

A nondirected network representing the electrical power grid of the western United States (*4*) showed an SRP with overrepresentation of tetrads 3 to 6 (Fig. 3). Nondirected networks of protein structure in which nodes are secondary-structure elements (α helices and β strands) and two nodes are connected if their distance is smaller than 10 Å have similar SRPs with overrepresented tetrads 3, 5, and 6. We compared these networks to model networks in which connections are deter-

mined on a lattice by geometrical proximity. In these geometrically constrained networks, the nodes are arrayed on a lattice (a line in one dimension, a plane in two dimensions, etc). Points that are closer than a distance R on the lattice are linked by an edge with probability p. Points at a distance greater than R are unlinked. The resulting subgraph distributions of these networks and their corresponding randomized versions can be analytically calculated. We find good agreement between the real-world protein structure and power-grid SRPs and the corresponding geometrical models with a similar number of nodes, edges, and clustering coefficient (Fig. 3).

A distinct family of SRPs was found for the Internet at the level of nondirected connections between autonomous systems (AS, which are groups of computers within which networking is handled locally, but between which data flows over the public Internet). We studied the structures of the AS network sampled at different time points from 1997 to 2001 (*12*). The SRP of the AS networks was similar despite their different sizes. We find that the SRP of these networks is very different from that of the geometrically constrained superfamily, with tetrads 2 to 4 underrepresented and tetrad 5 overrepresented.

Finally, we studied the preferential attachment model of Barabasi and Albert (BA) (*28*), which is widely used to study network evolution. In the BA model, a nondirected network is

grown node by node, connecting each new node to m existing ones. The probability of connecting to an existing node i increases with the number of edges it already has. We find that the SRP of these networks (Fig. 3) has different forms for $m = 1$, $m = 2$, and high m (*29*). This is because not all tetrad patterns can be created when $m = 1$ or 2. The present approach can thus be used to study model networks (*28*) and allow comparison of their local structure to that of real-world systems.

In the SOM Text, we also show the SRP of tetrads for the directed networks considered above. We find that generally tetrad profiles of related networks are similar. However, networks of different types in the same triad superfamily sometimes show distinct tetrad profiles, suggesting that higher order subgraph profiles can help refine network classification.

The present approach demonstrates that networks of the same type share not only network motifs, but also characteristic SPs with very similar proportions of motifs and antimotifs significance. In addition, we find several superfamilies of networks that have similar SPs even though they describe different systems of vastly different sizes. What do the superfamilies mean? One possibility is that the similarity in SP is accidental and that distinct evolutionary histories can lead to a similar local structure.

Another possibility is that the different systems in a superfamily have similar key circuit elements because they evolved to perform similar tasks. The latter possibility leads to intriguing hypotheses that connect networks from different disciplines. This can allow for a better understanding of a given network based on results from other networks in the same superfamily. It would be interesting to use the present approach to map the relation between the function and the local structure of real-world networks.

References and Notes

1. S. H. Strogatz, *Nature* **410**, 268 (2001).
2. M. Newman, *SIAM Rev.* **45**, 167 (2003).
3. A. L. Barabasi, R. Albert, *Science* **286**, 509 (1999).
4. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
5. S. Maslov, K. Sneppen, *Science* **296**, 910 (2002).
6. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
7. R. Milo *et al.*, *Science* **298**, 824 (2002).
8. N. Rosenfeld, M. B. Elowitz, U. Alon, *J. Mol. Biol.* **323**, 785 (2002).
9. S. Mangan, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11980 (2003).
10. S. Mangan, A. Zaslaver, U. Alon, *J. Mol. Biol.* **334**, 197 (2003).
11. The algorithm of (7) (available at www.weizmann.ac.il/mcb/UriAlon; see also http://arxiv.org/abs/cond-mat/0312028 and N. Kashtan *et al.*, *Bioinformatics*, in press) was used to generate randomized networks and to calculate the subgraph numbers in the real and randomized networks. The randomized networks preserved the number of incoming and outgoing edges for each node, as well as the number of mutual edges for each node (a mutual edge between x and y corresponds to x→y and y→x).
12. Networks and their sources are listed in table S1. Neural synapses of *C. elegans* included all data except those for muscle cells. Unlike (7), all connections were used, not only those with ≥5 synapses. Our definition of the TSP can be sensitive to rare events of a mutual edge within a network. We omitted mutual edges in networks where mutual edges make up <0.1% of the total number of edges (TRANSC-YEAST with one mutual edge, and TRANSC-YEAST-2 with two).
13. D. Thieffry, A. M. Huerta, E. Perez-Rueda, J. Collado-Vides, *Bioessays* **20**, 433 (1998).
14. T. Ishii, K. Yoshida, G. Terai, Y. Fujita, K. Nakai, *Nucleic Acids Res.* **29**, 278 (2001).
15. T. I. Lee *et al.*, *Science* **298**, 799 (2002).
16. N. Rosenfeld, U. Alon, *J. Mol. Biol.* **329**, 645 (2003).
17. G. Lahav *et al.*, *Nature Genet.* **36**, 147 (2004).
18. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
19. J. J. Tyson, K. C. Chen, B. Novak, *Curr. Opin. Cell Biol.* **15**, 221 (2003).
20. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
21. J. P. Eckmann, E. Moses, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5825 (2002).
22. M. A. J. Van Duijn, M. Huisman, F. N. Stokman, F. W. Wasseur, E. P. H. Zeggelink, *J. Math. Sociol.* **27** (2003).
23. P. Holland, S. Leinhardt, D. Heise, Eds., in *Sociological Methodology* (Jossey-Bass, San Francisco, 1975), pp. 1–45.
24. D. Cartwright, F. Harary, *Psychol. Rev.* **63**, 277 (1956).
25. R. F. Cancho, R. V. Sole, *Proc. R. Soc. London Ser. B.* **268**, 2261 (2001).
26. J. Allen, *Natural Language Understanding* (Pearson Addison Wesley, ed. 2, 1995).
27. The similarities seen in the TSPs of different networks can be further appreciated by inspecting the distribution of concentrations of subgraphs, $C_i = Nreal_i/(Nreal_i)$. Even though they show similar TSP profiles, networks of different types in the same superfamily can show quite different concentration distributions (SOM Text). The TSP can therefore highlight similarities that are not evident from merely counting subgraphs.
28. J. Berg, M. Lassig, *Phys. Rev. Lett.* **89**, 228701 (2002).
29. K. I. Goh, E. Oh, H. Jeong, B. Kahng, D. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12583 (2002).
30. We thank J. P. Eckmann, A. L. Barabasi, F. N. Stokman, M. A. J. Van Duijn, and the STKE, COSIN, GeNet, DBTBS, YPD, regulonDB, PDB database teams for valuable data sets, and M. E. J. Newmann, D. Chklovskii, M. Ronen, M. Lapidot, and I. Yanai for discussions. We thank NIH, Israel Science Foundation, and Minerva for support. R.M. is supported by a doctoral fellowship from the Horowitz Complexity Science Foundation.