Mining Bridge and Brick Motifs From Complex Biological Networks for Functionally and Statistically Significant Discovery

Chia-Ying Cheng, Chung-Yuan Huang, and Chuen-Tsai Sun, *Member, IEEE*

***Abstract*—A major task for postgenomic systems biology researchers is to systematically catalogue molecules and their interactions within living cells. Advancements in complex-network theory are being made toward uncovering organizing principles that govern cell formation and evolution, but we lack understanding of how molecules and their interactions determine how complex systems function. Molecular bridge motifs include isolated motifs that neither interact nor overlap with others, whereas brick motifs act as network foundations that play a central role in defining global topological organization. To emphasize their structural organizing and evolutionary characteristics, we define bridge motifs as consisting of weak links only and brick motifs as consisting of strong links only, then propose a method for performing two tasks simultaneously, which are as follows: 1) detecting global statistical features and local connection structures in biological networks and 2) locating functionally and statistically significant network motifs. To further understand the role of biological networks in system contexts, we examine functional and topological differences between bridge and brick motifs for predicting biological network behaviors and functions. After observing brick motif similarities between *E. coli* and *S. cerevisiae*, we note that bridge motifs differentiate *C. elegans* from *Drosophila* and sea urchin in three types of networks. Similarities (differences) in bridge and brick motifs imply similar (different) key circuit elements in the three organisms. We suggest that motif-content analyses can provide researchers with global and local data for real biological networks and assist in the search for either isolated or functionally and topologically overlapping motifs when investigating and comparing biological system functions and behaviors.**

***Index Terms*—Complex biological systems, network motif, network-oriented approach, strong/weak links.**

I. Introduction

NETWORK-ORIENTED approaches to complex biological systems are receiving significant attention in computational systems biology research. As a result, progress is being made toward defining the organizing principles governing the formation and evolution of complex social, biological, and technological networks. Statistical features in complex networks have been identified and investigated in biological systems, including the small-world property [2]–[6] of short paths between any two nodes and local clustering by nodes having multiple mutual neighbors [1], [3], [6]. The degree distributions of nodes in these natural networks frequently take on a long-tailed (power-law) shape—also referred to as a scale-free property—in which most nodes have only a few connections but a few have many connections [1], [7]–[9].

Hierarchical modularity [39] signatures that combine scale-free and local-clustering properties have recently been observed in metabolic, protein–protein interaction, and gene-regulatory networks [38], [43]. Beyond these global statistical features, a number of local structural motifs (building blocks) are providing insights to biological networks and revealing their own statistically significant patterns [7], [12]. These motifs consist of small subgraphs that occur in biological networks far more often than in randomized networks [7], [8], [10]–[15]. Some are thought to play important functional and information roles [7], [11], [15], [16].

Predicting network behaviors and functions requires the identification of functionally important and statistically significant motifs [1]–[3], [6]–[8], [15], [19], [20]. This is considered as a major challenge in computational biology. Instead of analyzing motif functions according to their real network topological features and effects [10], some researchers are analyzing networks in terms of their theoretical structures [7], [12], [15], [21], [22]. Using gene-transcription networks as an example, the strength of a factor's effect on a targeted gene's transcription rate is measured using an input function that is approximated via a logic [48] or Hill function [49]. Other researchers are using genetic algorithms to build complex networks in order to understand network functions and evolutionary mechanisms [50], [51]. While helpful in understanding network dynamic behaviors and functions, these simple frameworks seldom consider global-network features and local structures simultaneously [61]. Furthermore, motifs with statistical significance that are experimentally proven to be unimportant to network functions have no utility for investigating network functions and behaviors [8]. In this paper, we propose a method for locating functionally and statistically significant network motifs while simultaneously identifying global statistical features and local structural motifs in biological networks.

Manuscript received June 20, 2007; revised September 3, 2007. This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC96-2221-E-182-041 and in part by Chang Gung Memorial Hospital under Grants CMRPD260021 and CMRPD260031. This paper was recommended by Guest Editor F. Azuaje.

C.-Y. Cheng and C.-T. Sun are with the Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C.

C.-Y. Huang is with the Department of Computer Science and Information Engineering and the Research Center for Emerging Viral Infections, Chang Gung University, Taoyuan 333, Taiwan, R.O.C.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2007.908842

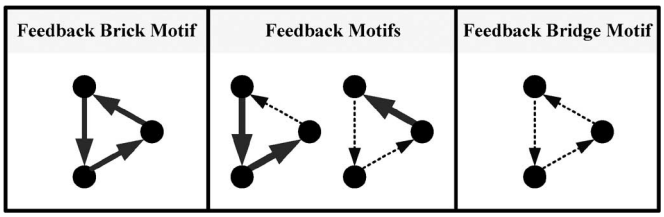
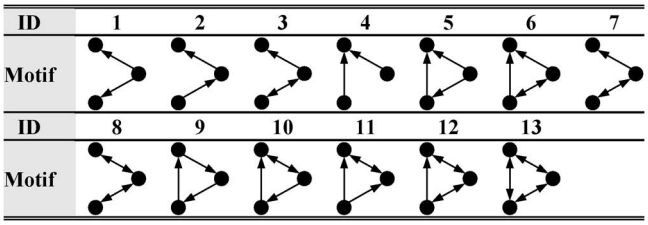


Fig. 1. Three-point feedback motifs can be divided into three categories. (Left) Feedback brick motif that is composed of three strong links. (Right) Feedback bridge motif that is composed of three weak links. (Middle) Other possible motifs were not considered because their topological properties are not specific. Note that the bridge and brick motifs shown here were extracted from a real complex network.

TABLE I
THIRTEEN THREE-NODE DIRECTED SUBGRAPHS WITHOUT WEIGHTED VALUES



When considering the global statistical features and local structural motifs of biological networks, it is worth noting that link properties (weights) exert strong impacts on network functions and dynamic behaviors [1], [17], [23], [24], [46]. Examples include the role of weak links in the "six degrees of separation" (i.e., small world) effect of interpersonal networks [23], [24] and the strength of predator–prey interactions that determine the stability of ecological communities [17]. Network researchers have reported that a weighted value representing interaction strength can be assigned to each link (edge) in a real network [1], [25]–[27]. Moreover, most genes and proteins do not have independent functions; instead, their roles are realized via complex interactions with other proteins, genes, and biomolecules [29], [40], [42], [43], [46]. We therefore considered network-motif-link strength [1], [25]–[27] in terms of two categories: bridge (consisting of weak links only) and brick motifs (consisting of strong links only) (Fig. 1, Table I). Serving as network foundations, brick motifs play a central role in defining global topological organization [42]; bridge motifs include isolated motifs that neither interact nor overlap with other motifs.

II. BACKGROUND

Most cellular functions are modular [34]–[36], with each module consisting of nodes that are physically or functionally connected in order to perform independent tasks. For example, invariant protein–protein and protein–ribonucleic acid (RNA) complexes (physical modules) at the core of many basic biological functions consist of temporally coregulated groups of molecules that govern various cell cycle stages [34]. Several methods in identifying topological and functional modules have been proposed, including network topological descriptions [25], [26], [39], [53], [59] and combining topologies with functional genomic data [38]. However, different methods tend to predict different boundaries between modules that are not sharply separated. This ambiguity is both a limitation of clustering methods and a consequence of a network's hierarchical modularity [38], [39].

It has been suggested that, in most situations, motifs serve as basic modular elements in complex networks [7], [12], [18], [37], [56]. Three additional pieces of evidence indicating that motifs have direct biological relevance are evolutionary motif conservation within yeast–protein interaction networks [38], [60], convergent evolution toward similar motif types in the transcriptional regulatory networks of various species [30], [38], and specific motif types that aggregate into large clusters in *E. coli* [42].

Milo *et al.*'s [7], [28] motif concept has served as the jumping-off point for many studies. For example, based on the notion that the number of distinct subgraphs (sets of nodes with specific sets of connecting edges) grows exponentially with the number of nodes, researchers have developed efficient and scalable heuristics in detecting specific subgraphs and in determining their frequencies in large networks [38], [52]. Some motif-identification methods have been used with stochastic networks containing intrinsic and/or experimental uncertainties [54], and a topological relationship between the large-scale attributes and local interaction patterns of complex networks has been reported [10], [52], [54]. Other researchers have focused on motif variations (including graphlets [55] consisting of small nonisomorphic network subgraphs) and motifs derived from families of similar (but not identical) patterns [14]. Note that subgraphs are not called motifs until they reach a specific statistical significance.

Attempts have been made to observe local network structures in order to determine network motif functions, with simulated dynamic systems being used to detect dynamic motif features and behaviors [12], [33], [45]. However, some dynamics remain undetected due to unknown and perhaps complex functional dependences between nodes, lack of knowledge of the parameters defining specific instances of motifs in real networks, or "unmodeled" interactions that may be absent from network representation but are nevertheless relevant [33].

Others are studying the importance of links, also called interactions [29], [42]. Individual research teams have reported the following: 1) crosstalk (meaning module connectivity that allows one function to influence another) is vital for cellular-function coordination [36], [43] and 2) the higher the number of internal connections in a motif, the more conserved the motif across species during evolution [40]. Shortcut links in small-world networks also play an important role, with some forming clusters that are found in many biological networks [29], [36], [40], [42]. As described in the following section, we adopted an approach that combines statistically significant building blocks with link properties to locate statistically and functionally significant network motifs in biological networks.

III. BRIDGE AND BRICK MOTIF DETECTION METHODS

To ensure that the concepts and methods described in this paper can be applied to any complex biological network, the link-weighted value $\mathrm{Link}(\nu, w)$ for any edge between nodes ν

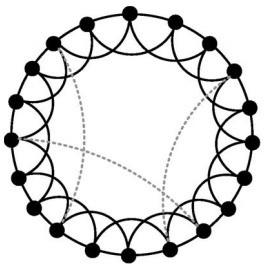


Fig. 2. Small-world model. Black signifies strong links and gray weak links (network shortcuts).

and w is expressed as its hypergeometric coefficient $C_{\nu,w}$ (1) [41]. This value (frequently used to measure cluster enrichment and cooccurrence significance) is expressed as

$$\begin{aligned}&\mathrm{Link}(\nu,w)\\&\quad=C_{V,W}\\&\quad=-\log\sum_{i=|N(\nu)\cap N(w)|}^{\min(|N(\nu)|,|N(w)|)}\frac{\binom{|N(\nu)|}{i}\times\binom{T-|N(\nu)|}{|N(w)|-i}}{\binom{T}{|N(w)|}}\end{aligned}\tag{1}$$

where $|N(x)|$ is the neighborhood size of node x, and T the total number of nodes in the biological network of interest. The summation in the hypergeometric coefficient $C_{\nu,w}$ can be represented as the probability in obtaining a number of mutual neighbors between nodes ν and w at or above the observed number when the neighborhoods are independent. The hypergeometric coefficient $C_{\nu,w}$ is consequently defined as the negative log of this summation. Given the neighborhood sizes of the ν and w nodes and the T total number of nodes in the biological network, the higher the value of $C_{\nu,w}$, the higher the number of overlapping neighbors between ν and w—an indication that $\mathrm{Link}(\nu,w)$ has a higher clustering coefficient. Otherwise, it does not belong to any specific cluster (Fig. 2).

The threshold Link_{AVG} that determines the strength of any link is the mean weighted value of all links in 1000 randomized networks. Each node in a randomized network has the same number of incoming edges (in-degrees) and outgoing edges (out-degrees) as its corresponding node in a real network. Furthermore, randomized networks preserve the same number of appearances of all $(n-1)$ node subgraphs as in the real (original) network [7]. When the weighted value $\mathrm{Link}(\nu,w)$ of a link between nodes ν and w in a randomized or real network is smaller than the threshold Link_{AVG} minus two standard deviations, the link is considered "weak"; all other links are considered "strong." Individual researchers can define criteria for strong and weak links according to their specific needs.

We expanded Milo *et al.*'s methodology [7] for identifying bridge and brick motifs in complex networks to include the following steps.

1) Calculate the weighted value of each link in a network of interest and an ensemble of random networks to calculate the significance of n-node subgraphs (1).
2) Label all weighted links in the network of interest and random network ensemble as strong or weak according to a benchmark of two standard deviations from the mean weighted value of all links in the ensemble. Links with weighted values below the benchmark are labeled as weak.
3) Identify all n-node bridge/brick subgraph types in the network of interest and random network ensemble.
4) Mark all n-node bridge/brick subgraph types by calculating their numbers in the network of interest and random network ensemble. An n-node bridge/brick subgraph type is selected as a representative motif only if its frequency in the network of interest far exceeds its frequency in the ensemble.

Motif frequency can be used to measure levels of similarity between two networks. Furthermore, it is possible to calculate Z_{score} for all bridge and brick motifs and to establish significance profiles (SPs) in any network by expanding Milo *et al.*'s [7], [28] methods. In (2), $Z_{\mathrm{score}}(\mathrm{Bridge}_i)$ represents the statistical significance of the ith kind of bridge motif in a network

$$Z_{\mathrm{Score}}(\mathrm{Bridge}_i)=\frac{N_{\mathrm{real}}(\mathrm{Bridge}_i)-\langle N_{\mathrm{random}}(\mathrm{Bridge}_i)\rangle}{STD\left(N_{\mathrm{random}}(\mathrm{Bridge}_i)\right)}\tag{2}$$

where $N_{\mathrm{real}}(\mathrm{Bridge}_i)$ represents the time of appearance of the ith type of bridge motif in a network, and $\langle N_{\mathrm{random}}(\mathrm{Bridge}_i)\rangle$ and $STD(N_{\mathrm{random}}(\mathrm{Bridge}_i))$, respectively, represent the mean and standard deviation of the time of appearance of the ith type of bridge motif in a randomized network ensemble. In (3), $SP(\mathrm{Bridge}_i)$ is the vector of $Z_{\mathrm{score}}(\mathrm{Bridge}_i)$ normalized to a length of one. The normalization emphasizes the relative rather than the absolute significance of the ith type of bridge motif. As shown in (4) and (5), $Z_{\mathrm{score}}(\mathrm{Brick}_i)$ and $SP(\mathrm{Brick}_i)$ can be derived in the same manner

$$SP(\mathrm{Bridge}_i)=\frac{Z_{\mathrm{Score}}(\mathrm{Bridge}_i)}{\left(\sum Z_{\mathrm{Score}}(\mathrm{Bridge}_i)^2\right)^{1/2}}\tag{3}$$

$$Z_{\mathrm{Score}}(\mathrm{Brick}_i)=\frac{N_{\mathrm{real}}(\mathrm{Brick}_i)-\langle N_{\mathrm{random}}(\mathrm{Brick}_i)\rangle}{STD\left(N_{\mathrm{random}}(\mathrm{Brick}_i)\right)}\tag{4}$$

$$SP(\mathrm{Brick}_i)=\frac{Z_{\mathrm{Score}}(\mathrm{Brick}_i)}{\left(\sum Z_{\mathrm{Score}}(\mathrm{Brick}_i)^2\right)^{1/2}}.\tag{5}$$

IV. Experimental Results

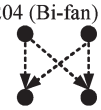
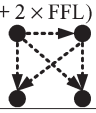
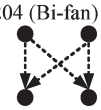
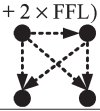
We applied our proposed method to *E. coli* (bacteria) and *S. cerevisiae* (yeast) transcriptional gene regulation networks [28]. Network and source data are listed in Tables II and III. Additional data are available at ftp://www.csie.cgu.edu.tw/.

In both networks, nodes represent operons (i.e., one or more genes transcribed on the same mRNA [28]), and directed links represent transcriptional regulatory relationships between operons that encode transcription factors (TFs) and operons regulated by TFs. In both *E. coli* and *S. cerevisiae*, we observed many v-out and feedforward-loop (FFL) brick motifs (ID = 1 and 5, respectively) (Table III). The FFL (a three-gene subgraph) is composed of two input TFs, one regulating the other

TABLE II
EDGE AND NODE DEFINITIONS, NETWORK SIZES, AND REFERENCES FOR FIVE GENE REGULATION NETWORKS

Network Type	Common Feature	Network	Nodes	Edges	Description
Gene Regulation (transcription)	Directed graph in which nodes represent genes and edges are directed between genes as regulated by a transcription factor.	*E. coli*	424	519	*Eschericahia coli* [12]
		S. cerevisiae (yeast)	688	1079	*Saccaromyces cerevisiae* [33]
		Drosophila	110	307	*Drosophila melanogaster* www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm
		Sea urchin (endomesoderm)	43	58	*Sea urchin* [33]
		C. elegans	280	2170	*C. elegans* (all synaptic connections used; not restricted to those with ≥ 5 synapses) [28]

TABLE III
BRICK AND BRIDGE MOTIFS IN FIVE GENE REGULATION NETWORKS

Network	Nodes	Links	Motif Type	ID	N_{Real}	N_{Random}±STD	Z_{Score}
E. coli	424	519	Brick	1	402	22.8±19.8	19.17
			Brick	5	42	7.3±3.2	10.72
			Bridge	204 (Bi-fan) 	107	46.8±15.0	4.01
			Bridge	206 (Bi-fan + 2 × FFL) 	51	2.5±3.2	15.26
S. cerevisiae (yeast)	688	1079	Brick	1	416	17.6±13.9	28.74
			Brick	5	70	13.6±4.0	13.98
			Bridge	204 (Bi-fan) 	1673	276.2±41.2	33.99
			Bridge	206 (Bi-fan + 2 × FFL) 	157	4.3±3.6	42.66
Drosophila	110	307	Brick	1	354	123.1±27.8	8.31
			Brick	2	264	108±23.6	6.61
			Brick	5	109	29±7.7	10.45
			Brick	11	14	2.2±1.6	7.55
Sea urchin (endomesoderm)	43	58	Brick	1	42	24.4±11.2	1.57
			Brick	6	5	1.4±1.4	2.52
C. elegans	280	2170	Bridge	5	740	489.6±32.9	7.60
			Bridge	6	141	53.6±8.7	10.05
			Bridge	11	213	59.1±8.9	17.22
			Bridge	12	75	24.6±4.8	10.41
			Brick	1	2479	950.5±85.1	17.97
			Brick	5	297	46.5±7.4	33.76
			Brick	11	31	2.8±1.8	15.96
			Brick	12	11	0.5±0.7	15.63

and both jointly regulating a target gene [13]. The observation that FFL bridge motifs do not exist in either network supports previous findings indicating that most motifs do not function in isolation but overlap with known biological functions [29], [42], [44], [56]. Specifically, one FFL motif cluster overlaps with the flagella-motor module, and another contains a significant number of elements responsible for regulating the *E. coli* aerobic/anaerobic switch [42]. We suggest that, since most FFL motifs consist of strong links, many (if not all) FFL motif interactions can be used as parts of other motifs or modules (e.g., flagella motor, osmoregulated porin gene, oxidative stress response, methionoine-biosynthesis modules) in a manner that makes the most efficient use of each gene or operon archive [42], [56]. Accordingly, FFL brick motifs are viewed as having an optimal design in terms of convergent evolution in transcriptional gene regulation networks [30].

The other motif type that is well represented in both networks is the four-gene bifan pattern associated with bridge motifs (Table III). The bifan consists of two input TFs, one never regulating the other but both jointly regulating two target genes. In *E. coli*, 208 of the 209 bifan motifs we observed combine to create dual-motif clusters, in which most links are shared by at least two adjacent motifs in addition to multiple nonadjacent motifs [42]. We did not find any bifan brick motifs but noted 107 bifan bridge motifs that do not overlap with other motifs, indicating that they function by themselves. From this, we

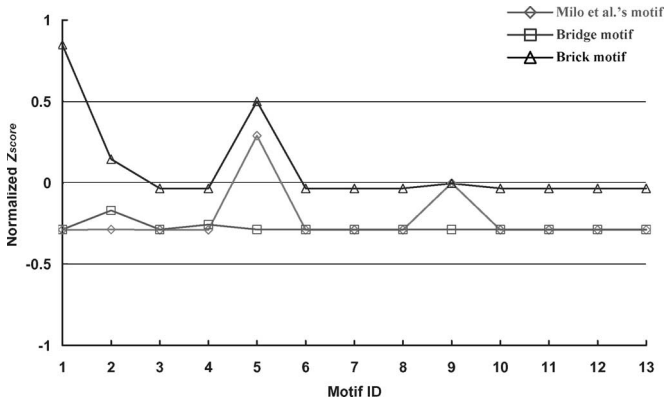


Fig. 3. Comparisons of TSPs for our bridge and brick motifs and Milo *et al.*'s [7], [28] *E. coli* motifs.

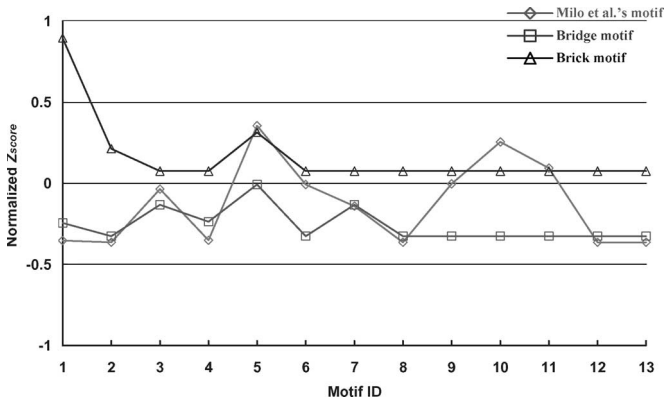


Fig. 4. Comparisons of TSPs for our bridge and brick motifs and Milo *et al.*'s [7], [28] *S. cerevisiae* (yeast) motifs.

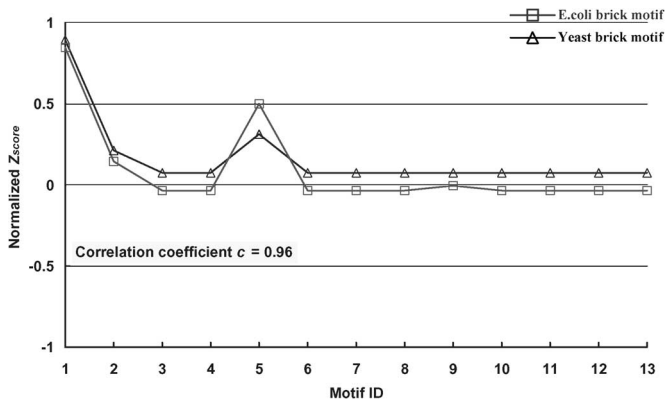


Fig. 5. Brick motif ratio profiles for two gene regulation networks: *E. coli* and *S. cerevisiae* (yeast).

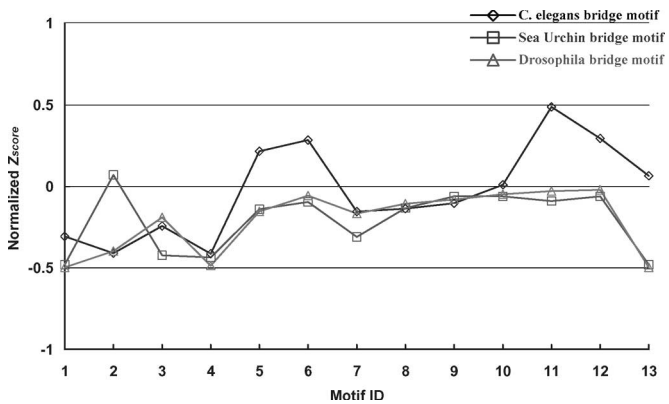


Fig. 6. Bridge motif ratio profiles for three gene regulation networks: *C. elegans*, sea urchin, and *Drosophila*.

inferred a low coregulation ratio for two operons, in which one regulates the other.

We will use the bifan bridge motif consisting of *aroL*, *mtr*, TrpR, and TyrR as an example. The combination of the TyrR protein and TrpR repressor are responsible for regulating other aromatic amino acid transport genes [57]. The TyrR protein plus either phenylalanine or tyrosine is responsible for *mtr*-gene activation, while a combination of the TrpR repressor plus tryptophan represses the *mtr* gene [58]. Both TyrR and TrpR regulate the expression of the *aroL* gene-encoding enzyme shikimate kinase II in *E. coli* [42]. We also found 51 brick motifs $(\mathrm{ID} = 206)$ consisting of combinations of FFL and bifan motifs. As Dobrin [56] reports, these motifs form a heterologous motif superstructure. Our results for *S. cerevisiae* are similar to those for *E. coli*. After comparing our results with Milo *et al.*'s results [28], we determined that v-out $(\mathrm{ID} = 1)$ and FFL brick motifs $(\mathrm{ID} = 5)$ play important roles in both networks (Figs. 3 and 4). Furthermore, the brick-motif ratio profiles in the two gene regulation networks are very similar (correlation coefficient $c = 0.96$) (Fig. 5), even though they contain relatively few brick motifs [28].

We made an effort to learn more about the relationship between coherent (incoherent) FFLs [12] and brick (bridge) FFLs. Since each of the three FFL interactions can be either activating or repressing, FFLs have eight possible structural types [13], [42]. The four incoherent FFL types act as sign-sensitive accelerators that shorten the response time of target-gene expression following stimuli in one direction (e.g., off to on) but not the other. The four coherent FFL types act as sign-sensitive delays. *E. coli* contains 34 coherent FFLs, eight incoherent FFLs [42], 29 brick-coherent FFLs, and six brick-incoherent FFLs. Accordingly, the difference in coherent (incoherent) FFL frequencies cannot be simply explained by the relative abundances of brick and bridge motifs in a network.

Next, we applied our proposed method to transcription networks that guide development in *Drosophila melanogaster* and sea urchin and synaptic wiring in *Caenorhabditis elegans* (Table II). As in the two gene regulation networks, brick TSPs were more significant than bridge TSPs in these three networks. However, we also determined that four bridge motifs (ID = 5, 6, 11, and 12) in *C. elegans* are very significant (Table III), indicating the greater presence of isolated motifs. We suggest that these bridge motifs constitute the main difference between the *C. elegans* network and the *Drosophila* and sea urchin networks (Fig. 6). Similarities (differences) in bridge and brick motifs imply similar (different) key circuit elements in each organism.

To validate the respective roles of weak and strong links, we removed equal percentages of each (as well as random links). We found that in *E. coli* and *S. cerevisiae*, the greater the number of strong links removed, the lower the clustering coefficient relative to the randomly removed links. In contrast, the greater

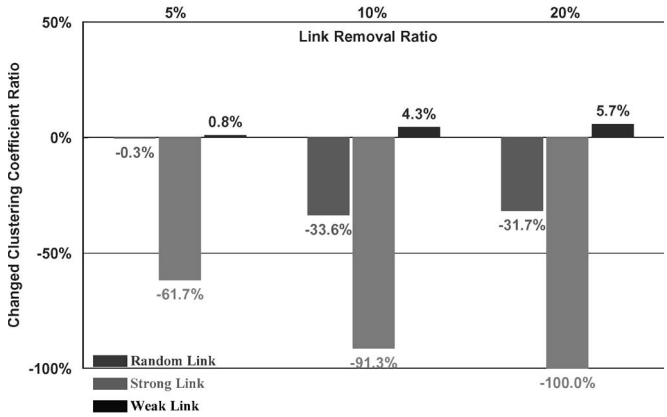


Fig. 7. Relationships between clustering coefficients and different removal ratios for three *E. coli* link types.

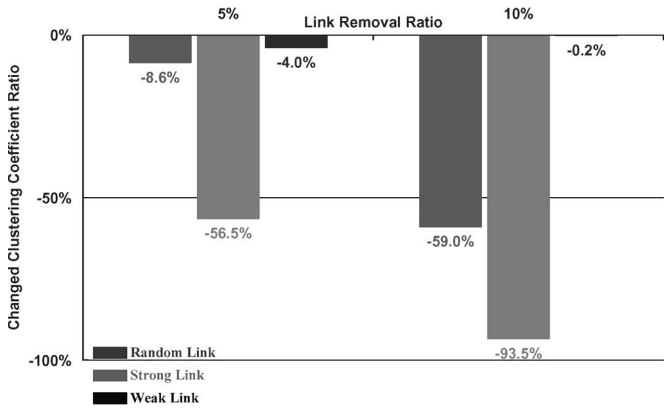


Fig. 8. Relationships between clustering coefficients and different removal ratios for three *S. cerevisiae* (yeast) link types.

the number of weak links removed, the higher the clustering coefficient relative to the randomly removed links (Figs. 7 and 8). Note that the average clustering coefficient increases when weak links are removed—i.e., when the clustering coefficient of a weak link's end node is calculated, its neighbors do not include the same link's other end node. The average coefficient increases after the weak links are removed, because the two end nodes do not share a large number of common neighbors. We did not compute the average degree of separation in the network after removing links, because a network may become broken and disconnected after a link is removed, and the definition of average degree of separation is based on a connected network. Note that our approach is insensitive to data errors; significant network motif sets in the two gene regulation networks do not change a great deal even when 40% of their edges are removed (Figs. 9 and 10). All altered results (red curves) shown in Figs. 9 and 10 represent average values for 30 runs. Our sensitivity-analysis results confirmed great similarities between the original and altered networks after randomly removing 40% of their links. According to the triad SP (TSP) of brick motifs, the original and altered networks belong to the same superfamily.

As shown in Fig. 11, the link weight distribution is extremely polarized (either zero or > 2), which supports our criterion for distinguishing between strong and weak links (i.e., mean weighted value $\text{Link}_{AVG} = 0.9$ and standard deviation $\text{Link}_{STD} = 0.04$ for all links in 1000 randomized networks). In most cases, random networks have many more weak than strong links. At least one researcher has suggested that high degree of clustering is a generic feature of biological networks [38].

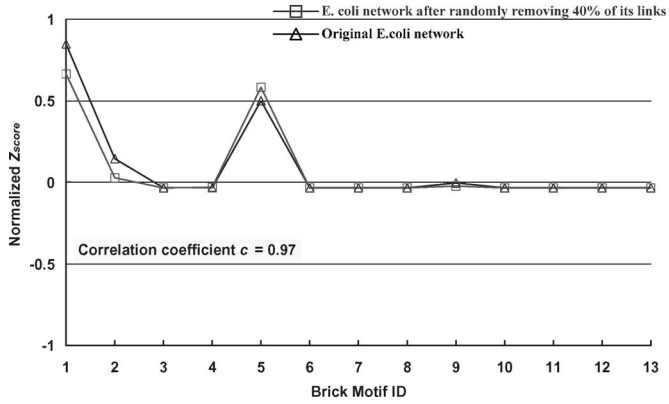


Fig. 9. Comparison of (blue curve) original and (red curve) altered brick motif ratio profiles for *E. coli* after randomly removing 40% of its links. Altered results represent average values for 30 runs.

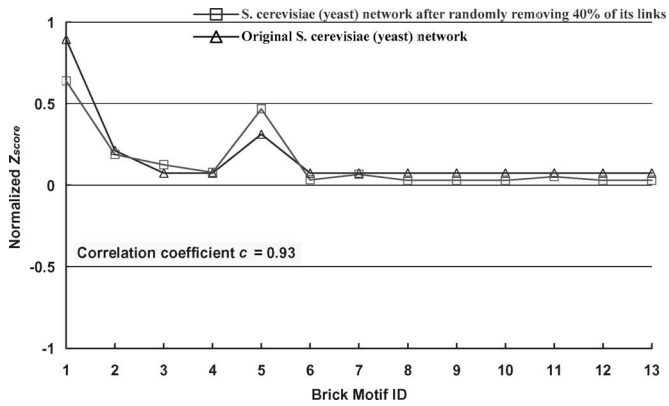


Fig. 10. Comparison of (blue curve) original and (red curve) altered brick motif ratio profiles for *S. cerevisiae* (yeast) after randomly removing 40% of its links. Altered results represent average values for 30 runs.

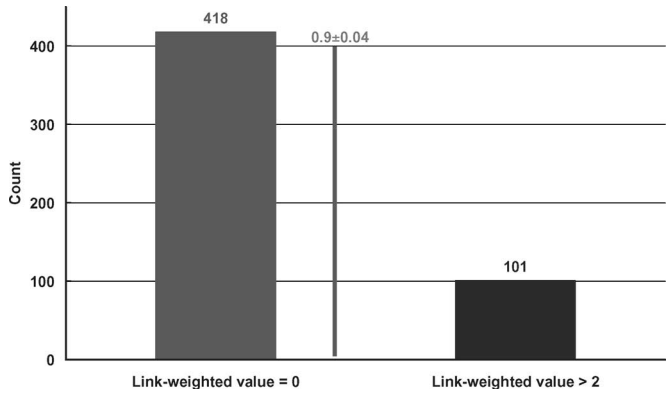


Fig. 11. Distribution of five link weights. Average mean and standard deviation of link weights for randomized networks were calculated as 0.90 ± 0.04.

The link property is a good indicator of cellular function robustness. The simplest strategy for protecting against the failure of a specific component is to provide alternative ways to perform that component's function. At the molecular level,

this backup strategy (or genetic buffering) [31] can be carried out by duplicate genes with identical roles or by different genes that constitute an alternate but functionally overlapping path [36]. Researchers can use brick motifs to explore identical genes that diverge functionally, reasons why the biological networks of unreliable elements still perform reliably [33], and the degeneracy phenomenon [32], [47].

V. Conclusion

According to our definitions of weighted links and network motifs and the results of our validation experiments using two gene transcription regulation networks, we conclude that the presence of bridge and brick motifs in a biological network is closely associated with network topological structures (particularly local connections) but not with network size (i.e., number of nodes). Bridge motifs can assist in the identification of isolated motifs, and brick motifs can be used to locate motifs whose functions overlap. This combination of a statistically significant motif and strong- or weak-link properties provides insight to the structural organizing principles and functions of networks. It can also serve as a method for analyzing biological system robustness.

References

[1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[2] M. E. J. Newman and D. J. Watts, "Scaling and percolation in the small-world network model," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 60, no. 6, pp. 7332–7342, Dec. 1999.

[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[4] R. F. I. Cancho, C. Janssen, and R. V. Sole, "Topology of technology graphs: Small world patterns in electronic circuits," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 4, p. 046 119, Oct. 2001.

[5] A. Barrat and M. Weigt, "On the properties of small-world network models," *Eur. Phys. J. B*, vol. 13, no. 3, pp. 547–560, Jan. 2000.

[6] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.

[7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, Oct. 2002.

[8] Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, "Comment on 'Network motifs: Simple building blocks of complex networks' and 'Superfamilies of evolved and designed networks'," *Science*, vol. 305, no. 5687, p. 1107, Aug. 2004.

[9] K. I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, "Classification of scale-free networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 99, no. 20, pp. 12 583–12 588, Oct. 2002.

[10] A. Vazquez, R. Dobrin, D. Sergi, J. P. Eckmann, Z. N. Oltvai, and A. L. Barabási, "The topological relationship between the large-scale attributes and local interaction patterns of complex networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 101, no. 52, pp. 17 940–17 945, Dec. 2004.

[11] H. S. Moon, J. Bhak, K. H. Lee, and D. Lee, "Architecture of basic building blocks in protein and domain structural interaction networks," *Bioinformatics*, vol. 21, no. 8, pp. 1479–1486, Apr. 2005.

[12] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nat. Genet.*, vol. 31, no. 1, pp. 64–68, May 2002.

[13] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 100, no. 21, pp. 11 980–11 985, Oct. 2003.

[14] J. Berg and M. Lassig, "Local graph alignment and motif search in biological networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 101, no. 41, pp. 14 689–14 694, Oct. 2004.

[15] M. Middendorf, E. Ziv, and C. H. Wiggins, "Inferring network mechanisms: The Drosophila melanogaster protein interaction network," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 102, no. 9, pp. 3192–3197, Mar. 2005.

[16] U. Alon, "Biological networks: The tinkerer as an engineer," *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[17] J. Bascompte, C. J. Melian, and E. Sala, "Interaction strength combinations and the overfishing of a marine food web," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 102, no. 15, pp. 5443–5447, Apr. 2005.

[18] J. J. Rice, A. Kershenbaum, and G. Stolovitzky, "Lasting impressions: Motifs in protein–protein maps may provide footprints of evolutionary events," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 102, no. 9, pp. 3173–3174, Mar. 2005.

[19] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, Dec. 2000.

[20] J. P. Eckmann and E. Moses, "Curvature of co-links uncovers hidden thematic layers in the World Wide Web," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 99, no. 9, pp. 5825–5829, Apr. 2002.

[21] S. C. Manrubia and J. F. Poyatos, "Motif selection in a model of evolving replicators: The role of surfaces and limited transport in network topology," *Europhys. Lett.*, vol. 64, no. 4, pp. 557–563, Nov. 2003.

[22] S. Mangan, A. Zaslaver, and U. Alon, "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks," *J. Mol. Biol.*, vol. 334, no. 2, pp. 197–204, Nov. 2003.

[23] M. Granovetter, "The strength of weak ties: A network theory revisited," *Sociol. Theory*, vol. 1, pp. 201–223, 1983.

[24] J. Davidsen, H. Ebel, and S. Bornholdt, "Emergence of a small world from local interactions: Modeling acquaintance networks," *Phys. Rev. Lett.*, vol. 88, no. 12, p. 128 701, Mar. 2002.

[25] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[26] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, p. 026 113, Feb. 2004.

[27] H. Zhu and Z. X. Huang, "Navigation in a small world with local information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 3, p. 036 117, Sep. 2004.

[28] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, "Superfamilies of evolved and designed networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, Mar. 2004.

[29] D. Zhou and R. Yang, "Global analysis of gene transcription regulation in prokaryotes," *Cell. Mol. Life Sci.*, vol. 63, no. 19/20, pp. 2260–2290, Oct. 2006.

[30] G. C. Conant and A. Wagner, "Convergent evolution of gene circuits," *Nat. Genet.*, vol. 34, no. 3, pp. 264–266, Jul. 2003.

[31] J. L. Hartman, IV, B. Garvik, and L. Hartwell, "Principles for the buffering of genetic variation," *Science*, vol. 291, no. 5506, pp. 1001–1004, Feb. 2001.

[32] G. M. Edelman and J. A. Gally, "Degeneracy and complexity in biological systems," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 98, no. 24, pp. 13 763–13 768, Nov. 2001.

[33] R. J. Prill, P. A. Iglesias, and A. Levchenko, "Dynamic properties of network motifs contribute to biological network organization," *PLoS Biol.*, vol. 3, no. 11, p. e343, Nov. 2005.

[34] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, no. 6761, pp. C47–C52, Dec. 1999.

[35] J. D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, "Evidence for dynamically organized modularity in the yeast protein–protein interaction network," *Nature*, vol. 430, no. 6995, pp. 88–93, Jul. 2004.

[36] J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle, III, and J. Doyle, "Robustness of cellular functions," *Cell*, vol. 118, no. 6, pp. 675–685, Sep. 2004.

[37] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, and H. Margalit, "Network motifs in integrated cellular networks of transcription-regulation and protein–protein interaction," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 101, no. 16, pp. 5934–5939, Apr. 2004.

[38] A. L. Barabási and Z. N. Oltvai, "Network biology: Understanding the cell's functional organization," *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, Feb. 2004.

[39] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.

[40] A. Vespignani, "Evolution thinks modular," *Nat. Genet.*, vol. 35, no. 2, pp. 118–119, Oct. 2003.

[41] D. S. Goldberg and F. P. Roth, "Assessing experimentally derived interactions in a small world," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 100, no. 8, pp. 4372–4376, Apr. 2003.

[42] R. Dobrin, Q. K. Beg, A. L. Barabási, and Z. N. Oltvai, "Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network," *BMC Bioinformatics*, vol. 5, p. 10, 2004.
[43] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, May 2002.
[44] H. W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer, and A. P. Zeng, "An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs," *Nucleic Acids Res.*, vol. 32, no. 22, pp. 6643–6649, Dec. 2004.
[45] P. J. Ingram, M. P. H. Stumpf, and J. Stark, "Network motifs: Structure does not determine function," *BMC Genomics*, vol. 7, p. 108, 2006.
[46] S. Wuchty, A. L. Barabási, and M. T. Ferdig, "Stable evolutionary signal in a yeast protein interaction network," *BMC Evol. Biol.*, vol. 6, p. 8, 2006.
[47] D. C. Krakauer and J. B. Plotkin, "Redundancy, antiredundancy, and the robustness of genomes," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 99, no. 3, pp. 1405–1409, Feb. 2002.
[48] L. Glass and S. A. Kauffman, "The logical analysis of continuous, nonlinear biochemical control networks," *J. Theor. Biol.*, vol. 39, no. 1, pp. 103–129, Apr. 1973.
[49] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London, U.K.: CRC, 2007.
[50] S. A. Teichmann and M. M. Babu, "Gene regulatory network growth by duplication," *Nat. Genet.*, vol. 36, no. 5, pp. 492–496, May 2004.
[51] N. Kashtan and U. Alon, "Spontaneous evolution of modularity and network motifs," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 102, no. 39, pp. 13 773–13 778, Sep. 2005.
[52] T. Aittokallio and B. Schwikowski, "Graph-based methods for analysing networks in cell biology," *Brief. Bioinf.*, vol. 7, no. 3, pp. 243–255, 2006.
[53] F. Luo, Y. Yang, C. F. Chen, R. Chang, J. Zhou, and R. H. Scheuermann, "Modular organization of protein interaction networks," *Bioinformatics*, vol. 23, no. 2, pp. 207–214, Jan. 2007.
[54] R. Jiang, Z. Tu, T. Chen, and F. Sun, "Network motif identification in stochastic networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 103, no. 25, pp. 9404–9409, Jun. 2006.
[55] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, pp. e177–e183, Jan. 2007.
[56] H. W. Ma, J. Buer, and A. P. Zeng, "Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach," *BMC Bioinformatics*, vol. 5, p. 199, 2004.
[57] S. Zhao, Q. Zhu, and R. L. Somerville, "The $\sigma(70)$ transcription factor TyrR has zinc-stimulated phosphatase activity that is inhibited by ATP and tyrosine," *J. Bacteriol.*, vol. 182, no. 4, pp. 1053–1061, Feb. 2000.
[58] J. P. Sarsero, P. J. Wookey, and A. J. Pittard, "Regulation of expression of the Escherichia coli k-12 mtr gene by TyrR protein and Trp repressor," *J. Bacteriol.*, vol. 173, no. 13, pp. 4133–4143, Jul. 1991.
[59] L. D. F. Costa, M. Kaiser, and C. C. Hilgetag, "Predicting the connectivity of primate cortical networks from topological and spatial node properties," *BMC Syst. Biol.*, vol. 1, p. 16, 2007.
[60] S. Wuchty, Z. N. Oltvai, and A. L. Barabási, "Evolutionary conservation of motif constituents within the yeast protein interaction network," *Nat. Genet.*, vol. 35, no. 2, pp. 176–179, Oct. 2003.
[61] J. Hallinan, "Gene duplication and hierarchical modularity in intracellular interaction networks," *Biosystems*, vol. 74, no. 1–3, pp. 51–62, Apr.–Jun. 2004.

Chia-Ying Cheng was born in Taiwan, R.O.C., in 1981. He received the B.S. and M.S. degrees in computer and information science from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2003 and 2005, respectively, where he is currently working toward the Ph.D. degree in computer science.

His current research interests include complex networks and systems, social simulations, and computational systems biology.

Chung-Yuan Huang was born in Taiwan, R.O.C., in 1970. He received the M.S. degree in computer information and science and the Ph.D. degree in computer science from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2000 and 2005, respectively.

He is currently an Assistant Professor with the Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan, Taiwan, where he is also a member with the Research Center for Emerging Viral Infections. His research interests include complex adaptive networks and systems, agent-based modeling and simulation for social science research, and computational epidemiology.

Chuen-Tsai Sun (S'90–M'92) received the B.S. degree in electrical engineering and the M.A. degree in history from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1979 and 1984, respectively, and the Ph.D. degree in computer science from the University of California, Berkeley, in 1992.

In 1991 and 1992, he participated in fuzzy-neural-network research with the Lawrence Livermore National Laboratory, Livermore, CA. Since 1992, he has been with the National Chiao Tung University, Hsinchu, Taiwan, where he is currently a Professor with the Department of Computer Science. His research interests include creative evolutionary systems, web-based collaborative learning, and computer simulation for social science research.