addition, patterns that represent similarity across edges are also discovered. Similarity and dissimilarity patterns are treated separately. This allows us to maintain the expression power of patterns we developed earlier [17] and yet to provide a larger set of patterns to the user. BISON is divided into a graphical visualization part and a text database interface for analysis and search. The network visualization component makes use of the Java Universal Network/Graph (JUNG) Framework [37]. The source code for BISON, documentation, and the pattern library are contained in the supplemental files.

Sources for network data

Network data were obtained from public databases, as well as experiments of our own and other researchers. A summary is given in Table 3. Four sources are described as follows: RegulonDB [2,47] provides 2,537 interactions defined by 142 regulators and 1,059 targets. Microarray data for all of the *E. coli* two-component systems [48] contain 1,028 interactions between 40 regulators and 372 targets. A previous compilation of data [33] contains 1,969 interactions over 26 regulators and 856 targets. Finally, a previous microarray experiment from our own laboratory [28] contributed 886 interactions over 2 regulators and 444 targets. The total number of interactions is 6,227 between 186 regulators and 1,934 targets. These sources were formatted for use in BISON. Since we wanted to link network data to node property information, we resolved all gene names to a common identifier (Blattner IDs [49]). Gene names that could not be linked to Blattner IDs were excluded from the network since they could not contribute to useful patterns.

Sources for property data

Property data were obtained from public sources, supplemented by our own set of properties. A summary is given in Table 4. The initial property data were obtained from the *E. coli* genome project [49,50] that includes 4,285 proteins and contains Gene Ontology (GO) classifications. The project yielded a total of 62 GO annotations (prefix "GO" in BISON). To extend these data, we searched the protein sequences for protein family domains. First, we gathered default data from the Pfam site [51,52] from which we extracted 1,032 annotations covering 2,271 proteins (prefix "PF" in BISON). In addition, we used the HMMER profile hidden Markov model software [53,54] that is used by Pfam to identify potential domain annotations using different criteria. Sequences from the *E. coli* Genome Project [49,50] were tested with the Pfam A domain models. Domains were detected at a cut-off e-value of 1e-10. From this software setup, we extracted 1,747 annotations covering 3,124 proteins (prefix "hmm"). Note that some proteins received multiple HMM annotations and some annotations overlap with the default Pfam annotations.

Data input files

All input data are collected in a single data directory (default_data). The directory includes a configuration file (bison.config) read by BISON that specifies names of data files. Files required by BISON: An entity file (ecoli_entity.txt) lists the nodes of the network and the set of properties available for each node. Specific node IDs (Blattner IDs) are gathered from this file. An alias file (ecoli_alias.txt) specifies the default gene names for the nodes. A synonym file (ecoli_syn.txt) lists additional names for the nodes. Finally, a pattern file (patterns.out) stores the patterns of entities and annotations discovered in the network. This file is generated by the pattern mining engine. BISON can accept results from other pattern generation libraries, provided the format requirements of the pattern output file are satisfied. The configuration file also contains a list of external web links that are used in combination with protein IDs to construct hyperlinks in the object information page.

Four network files (*.net) contain data from the four different sources we used (Table 3). Additional network files can be added to allow the user to integrate their own data and understand them in the context of the existing network. A line will have to be added in the bison.config file that lists the name of the new data file. Similarly, protein property information can be supplemented. If additional data are to be included in the pattern discovery, then the pattern mining script has to be re-run with the new data. This will result in a new pattern file (see above). Instructions for the addition of user-specific files are contained in the User Manual.

Table 3: Network data

Source	Interactions	Regulators	Regulated genes	Reference	File name
RegulonDB	2,537	142	1,059	[2,47]	regulon.net
Two-comp.	1,028	40	372	[48]	2component.net
Compilation	1,969	26	856	[33]	pruess.net
FlhD/FlhC	896	2	444	[28]	flhD_microarray.net
Total	**6,227**	**186**	**1,934**		