 doi:10.1093/bib/bbn020

VisANT: an integrative framework for networks in systems biology

Zhenjun Hu, Evan S. Snitkin and Charles DeLisi

Submitted: 6th February 2008; Received (in revised form): 23rd March 2008

Abstract

The essence of a living cell is adaptation to a changing environment, and a central goal of modern cell biology is to understand adaptive change under normal and pathological conditions. Because the number of components is large, and processes and conditions are many, visual tools are useful in providing an overview of relations that would otherwise be far more difficult to assimilate. Historically, representations were static pictures, with genes and proteins represented as nodes, and known or inferred correlations between them (links) represented by various kinds of lines. The modern challenge is to capture functional hierarchies and adaptation to environmental change, and to discover pathways and processes embedded in known data, but not currently recognizable. Among the tools being developed to meet this challenge is VisANT (freely available at http://visant.bu.edu) which integrates, mines and displays hierarchical information. Challenges to integrating modeling (discrete or continuous) and simulation capabilities into such visual mining software are briefly discussed.

Keywords: *network; integration; systems biology; metagraph; visualization*

INTRODUCTION

The ability to understand the cell as the complex system of spatially organized interacting molecules that it is, requires integrating large amounts of information [1, 2] of many different types and from many different spatial and temporal scales [1, 3–8]. It is a major challenge to represent this complex biological reality in a manner that allows for simple and intuitive visualization, and yet preserves the different levels of complexity in a way that allows for meaningful biological insights to be gained. A number of visualization tools have been developed [9–18] which have a goal representing large amounts of data in a manner that conveys functional information that would not otherwise be apparent. One of the earliest, VisANT [19, 20], now supports biological network mining and analysis, meaning that it is able to find previously unidentified networks, annotate them and display their hierarchical organization [21–23]. VisANT allows for targeted visualization of specific biological networks via features which allow for the querying of interactions for a given set of proteins/genes using names, accession numbers or gene IDs. VisANT also allows for the analysis of user-defined networks by providing several means for the creation/import of networks using basic data types, such as node/edge lists, PSI-MI [24] and BioPAX. Imported interactions and networks define a network 'workspace', which can be annotated and saved for printing or shared among others in the community in the format of VisML: an XML-based interpreting language and exchange format for storing annotation and visual layout of complex networks. In addition, networks and their statistical characteristics, such as degree distributions, can also be saved as JPEG images,

Corresponding author. Charles DeLisi, Bioinformatics Program, Boston University, 24 Cummington Street, Boston, MA02115, USA. E-mail: DeLisi@bu.edu

Zhenjun Hu is the assistant research professor at the Biomolecular Systems Laboratory in Boston University. His research interests are in the area of integrative, systems and network biology, with special interest in the development network related scientific software.
Evan Snitkin is a PhD student in the Bioinformatics Program at Boston University. His current research focuses on understanding metabolic function at a systems level. He has special interest in how metabolic networks respond to perturbations and how this relates to the organizing principles of metabolism.
Charles DeLisi is the Arthur G B Metcalf Professor of Science and Engineering, as well as the founding Chair of the all-University PhD Program in Bioinformatics at Boston University. His research focuses on the development and application of experimental and computational methods for high-throughput genomic and proteomic analysis to fundamental and applied questions in cell biology.

scalable vector graphics (SVG) or tab-delimited data files.

The rapid accumulation of diverse biological data, driven by the continuing development of high-throughput technologies, offers opportunities to gain insights by means of data integration, which exceed the insights reachable through the analysis of any single data type in isolation. We therefore view integration of data generated by various platforms into interpretable graphical formats, as indispensable to any network mining and visualization system. Achieving this goal, however, requires meeting a number of major challenges. These include the need for computational tools to (i) integrate disparate data types, both experimental and computational; (ii) navigate in an efficient manner enormous networks which are the consequence of global data integration and (iii) infer, mine and visually represent networks and the functional modules [7] of which they are comprised. Here we present the integration-related concepts, design principles and functions implemented in VisANT, especially the technology used to integrate multiple levels of biological knowledge into a single network.

DESIGN PRINCIPLES AND BASIC INTEGRATION CONCEPTS

Design principles

The VisANT system is designed based on the assumption that the majority of users will be scientists who have little programming skill. To this end, the principle 'easy to use' has taken the highest priority. VisANT takes two steps toward this goal. First, it has been kept compact (only 302 KB for the latest version 3.17), making it suitable as a web-based application to facilitate fast downloading, and to ease potentially troublesome software installation and upgrading; second, and probably more importantly, unlike other network visualization tools, such as Cytoscape [25], VisANT provides fundamental data integration services driven by the Predictome database [26], such as name resolution, which greatly lessens the burden of integrating data from various sources. For example, expression data can automatically be mapped to the appropriate nodes in a network once the *Name Normalization* function (see below) has been applied to its nodes, regardless of the different naming conventions used in expression and network data, respectively.

Database for the integration

The Predictome database was originally developed to store large numbers of relations amongst genes derived from evolutionary context-based methods, which exploit conservation and functionally driven changes [27] in genomic organization. Its function has been extended to serve as a data warehouse for the minimal data required for integration services. ~~Because integration is, can be time intensive, data providers often provide integrated data sets.~~ The recent development of the STRING database [28] is a good example of this trend, although its HTML-based representation schema has limited its application in network construction and visual analysis. To reduce maintenance costs of such data warehouse, we have developed scripts to automatically download and integrate various databases including KEGG [29], for which XML-like versions are available, GO [30], MIPS [31], BIND [32] and the human protein reference database [33]. VisANT accesses these data via the VisANT server using HTTP protocol.

Data integration and methodological integration

VisANT allows for knowledge integration in several different ways, which we classify under two rubrics: data integration (DI) and methodological integration (MI) [34, 35]. DI refers to the construction of various kinds of cell networks by transparently querying data from multiple sources. We distinguish two kinds of data integration: heavy and light. The former entails queries on locally stored data in the Predictome database. The latter refers to the ability to access, for individual or groups of nodes/edges, relevant references such as those in GenBank (sequence) [36] and OMIM (disease) [37] via options in a pull down menu.

DI is mainly a software engineering problem. It increases the number and kinds of relations, and to some extent confidence, when several sources of evidence all indicate the same relation, but it is not quantitative (Fig. 1IV). MI refers to the use of mathematical methods, such as one or another Bayesian variant or support vector machines, to carry out a principled and systematic integration. MI methods typically take as input weighted interactions of diverse types, and return confidence levels for inferred relations. MI is essential for drawing network inferences, and knowing how likely an inference is to be valid. Furthermore, MI methods can greatly reduce the complexity of an integrated

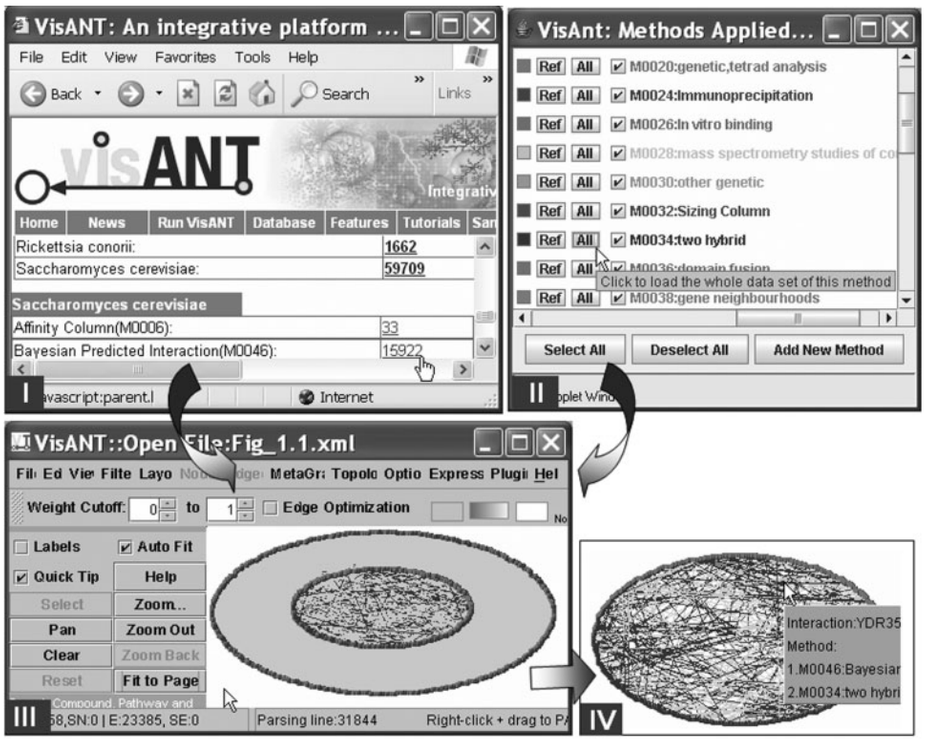

Figure 1: Method-based loading & filtering of large-scale interaction data set. (**I**) Directly load interactions associated with method M0046 into VisANT through interaction statistics page. (**II**) Directly load interactions associated with method M0034 through methods table in VisANT. (**III**) The combined interaction network of M0034 and M0046 shown in VisANT. (**IV**) The interaction network with the overlap of the two data sets created using built-in filter.

network by condensing diverse biological data into single all-encompassing relation.

NAME RESOLUTION

The first task in network integration is the unification of equivalent nodes which have been labeled using different classifications, such that total knowledge available for each node can be determined. This very basic, yet vital, unification operation is nontrivial in practice. In VisANT, unification is achieved using the *Name Normalization* function. Currently, this function can resolve the identities of genes or proteins, but not compounds.

Unification requires knowing whether the node refers to a gene or a protein. Although names of genes and proteins are often used interchangeably, gene IDs cannot reliably be used to represent proteins, primarily because splice variants exist. The *Name Normalization* function in VisANT is gene based, meaning that two protein nodes will be recognized as a single protein, if the same gene encodes them. This decision not to distinguish gene nodes from protein nodes was based primarily on the following observations:

- Independently created interaction databases often use different naming systems for proteins/genes. For example, MINT [12] uses a protein's UniProt [38] id in protein–protein interaction for *Homo sapiens* which, however, is represented using gene's HGNC (HUGO Gene Nomenclature Committee) [39] id in Biogrid [40].
- Genetic associations very often need to be compared to protein–protein interactions [41–43].
- High-throughput gene expression results must often be mapped onto a protein–protein interaction network or pathway.

Because genes often encode multiple proteins, using gene id to represent both gene and protein is a natural choice; otherwise, uncertainty may arise when integrating interaction data. For example, for a gene with two splice variants, the variant to use when integrating interactions between the MINT (protein-based) and Biogrid databases (gene-based)

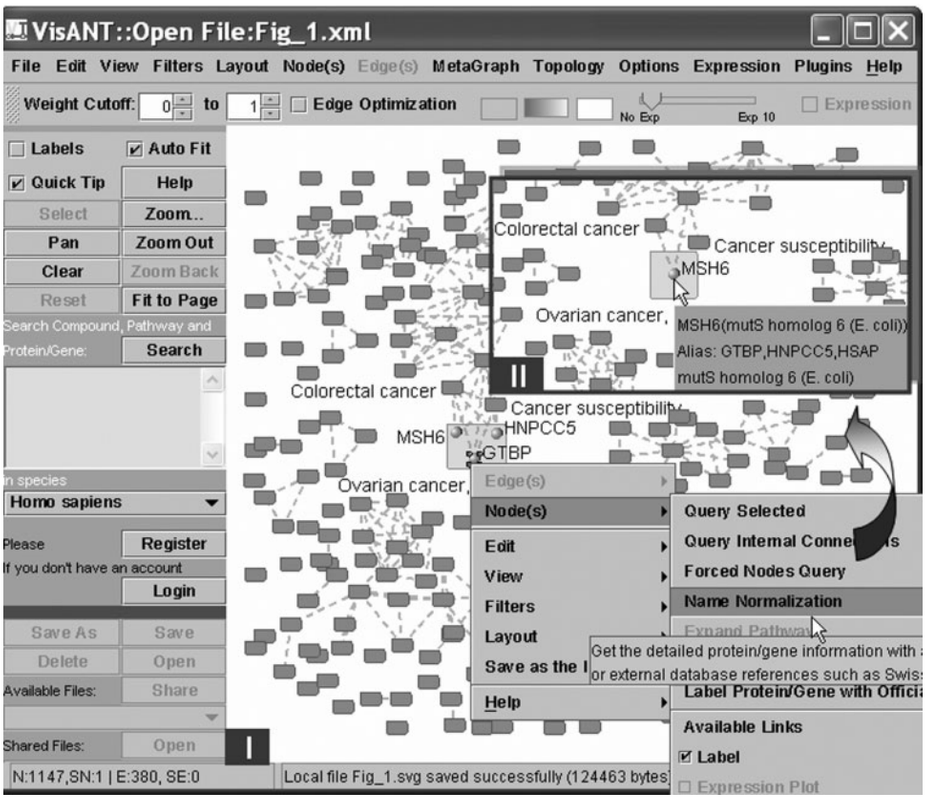

Figure 2: Network of cancers rebuilt in VisANT using metagraph with subset data of cancer extracted from the work of Goh and coworkers [47]. Each metanode (gray box) represents one type of cancer. The correlations between cancers are evaluated based on the number of shared genes. Mouse clicking, a metanode will reveal its substructure: genes involved in the cancer and their correlations to one another if any. An example of an expanded node for ovarian and endometrial cancers is shown. The original data shows that ovarian cancer of endometrial type involves three different genes (MSH6, GTBP, HNPCC5) (**I**) which are actually all the same gene with official name MSH6 as discovered by the *Name Normalization* function (**II**).

will be ambiguous. The shortcoming of this solution is that certain splice variants may only share a subset of interactions. This dilemma can be overcome if alternative splicing knowledge is represented using a metanode, as discussed in the section on future developments.

The data used for name resolution services are compiled from the Entrez Gene [36] and UniProt databases, as well as some organism-specific databases such as SGD [44], Flybase [45] and Wormbase [46]. Careful attention has been paid to assure that every ID/name will be mapped to only one gene for a given species.

Figure 2 shows an example of the use of the *Name Normalization* function. The original data [47] shows what is actually a single gene, but displayed as three different nodes each with its own label. This ambiguity is resolved by name normalization, as indicated in Figure 2. Subsequent to node unification, the gene aliases and a brief functional annotation will be shown as node's tooltip (Figure 2II), and HTTP links to the related nucleotide/protein sequences will be available in the corresponding menus. In addition, VisANT provides a special function to label a gene with its official name (Figure 2). This function will automatically be carried out when a gene's interactions are queried against the Predictome database. VisANT users can import their own mapping data using the ID-Mapping format, which is detailed in the VisANT user manual.

INTERACTION INTEGRATION

Large-scale studies have resulted in networks composed of various biological interaction types, such as protein–protein interactions, genetic interactions and transcriptional coexpression. A systems-level

understanding can be better achieved by integrating these diverse data types, such that not only the molecular assemblies involved can be deciphered, but also their functional connections. In VisANT, all interactions, regardless of whether they result from DI or MI, are classified based on the experimental/computational methods that report them. The name of the method is compatible with those defined in PSI-MI standard. Different types of interactions can be distinguished using their method names. For example, interactions based on two hybrid experiments are physical, while those with synthetic lethal are genetic, and donot necessarily imply a direct physical interaction. As shown in Figure 1, edges for the interactions between proteins are colored based on the interaction method, and edges resulting from DI are often associated with weights which can be used to quickly filter networks with different confidence scores (Figure 1III, weight cutoff).

The Predictome database stores relations based on some 57 different methods for 104 species. The interaction statistics on the VisANT web site lists the total number of interactions for each species, the methods contributing to these interaction totals, and the total number of interactions classified under each method. As shown in Figure 1, VisANT supports method-based fast loading of large numbers of interactions that can be obtained from the interaction statistics page or method table. The latter also enables method-based filtering of the interactions using a simple checkbox. The methods table also allows users to add, delete or customize the methods.

Among other issues illustrated in Figure 1, are two major challenges which need to be addressed in the implementation of any tool for supporting the visualization of large-scale interaction networks, namely, readability and performance. In general, a network becomes difficult to interpret with hundreds of nodes and edges; and for the majority of available tools, performance starts to deteriorate quickly when working with networks containing thousands of nodes and edges. Although VisANT has been highly optimized and can handle much larger networks (tested with 226k+ edges and nodes in a PC with 1 GB memory and 2.33 GHz CPU), some functions, such as spring force-based layout, can become very slow when working with larger networks. In the next section, we discuss metagraph-based technology, which addresses this bottleneck in network integration. And in the section on Future Developments, we discuss ways to provide additional support for computationally heavy tasks.

INTEGRATION OF MULTI-SCALE INFORMATION USING METAGRAPHS

The display of large-scale networks presents challenges relevant to the disciplines of graph theory, visualization and user-interface design. What is needed is a way to comprehensively view the entire network, while at the same time being able to limit the view to specific subgraphs, hiding or collapsing the rest of the network when it is not relevant.

As we discussed previously [21], hierarchical visualization is a suitable solution to the problem of integrating multiscale information, as it not only allows for complete network visualization, but also improves performance (Figure 3). However, a limitation of these hierarchical visualization technologies is that they are all based on compound graphs, which cannot display the intersection between two groups of nodes. This causes difficulties when attempting to accurately visualize biological networks, as it is very common for a gene/protein to participate in multiple cellular processes.

To circumvent this problem we developed a new type of graph—the metagraph [21] that allows a single node to have multiple visual instances through a node duplication function. The intersection between groups is achieved by creating two instances of the same node. As shown in Figure 4, node duplication also plays an important role in representing condition-dependency and combinatory control logic. In addition, metagraphs allow for an arbitrary number of hierarchical branches, such that multidimensional information can be integrated into a single network. For example, a coexpression module may be represented along with a GO functional module in the same network. Furthermore, metanodes can be nested indefinitely without restriction, providing the capacity for a complete description of the hierarchical nature of cellular function (Figure 3). Correlations between metanodes are usually inferred based on the properties of the subgraphs they encompass [21], and are represented as metaedges. Metaedges are dynamic; they are only present when both metanodes are collapsed. This metagraph capability allows VisANT users to reversibly and readily represent any set of related nodes [co-regulated genes, proteins in the same complex (Figure 3), genes involved in the same

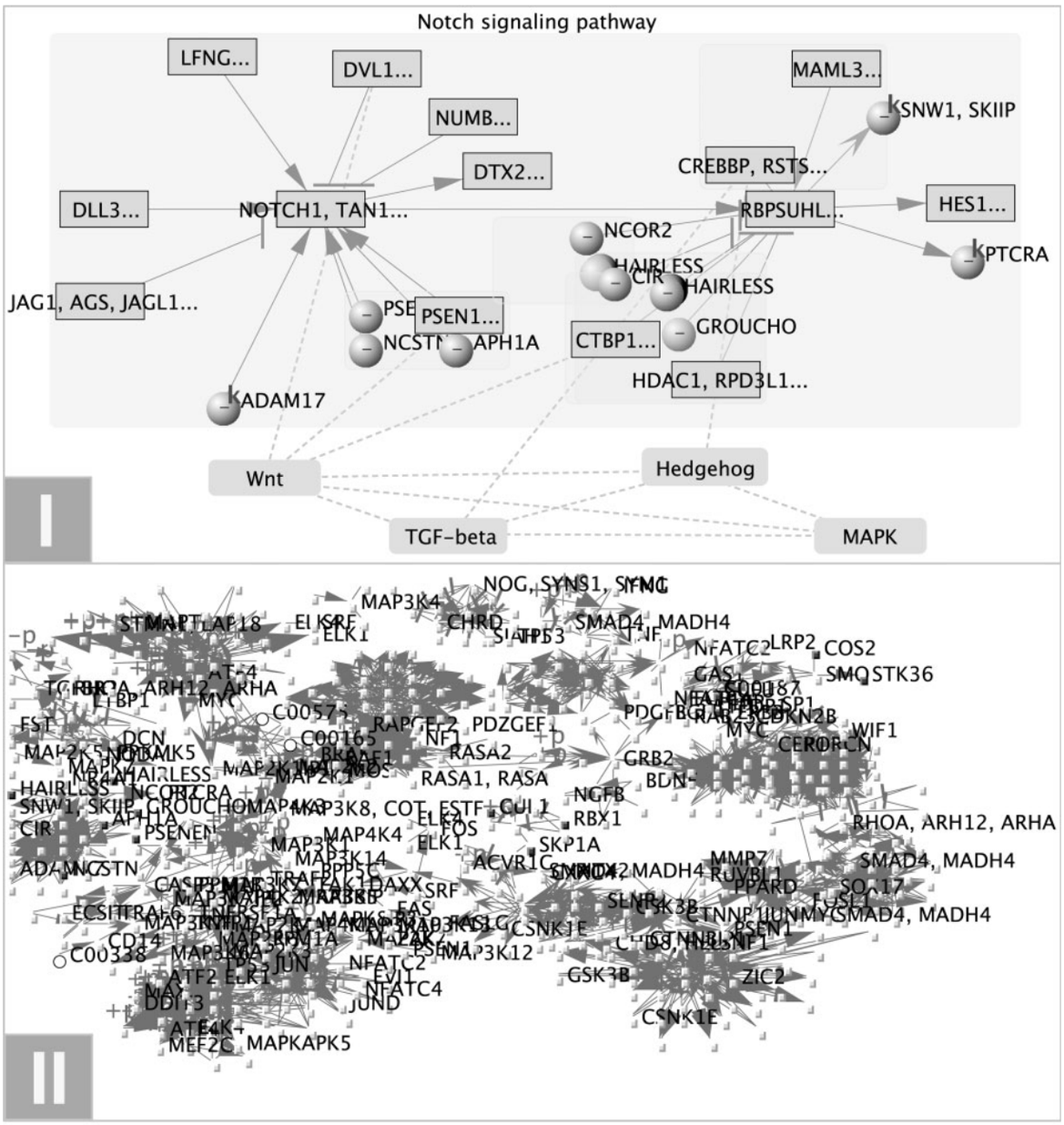

Figure 3: Improved readability and performance with multi-scale information integrated in pathway visualization using metagraph. Dark boxes represent the KEGG pathways; light boxes with dark border are contracted metanodes representing a group of proteins; gray boxes with light border representing the protein complex, filled dark circles represent protein and open circles represent compounds. (**I**) Five signaling pathways of *Homo sapiens* visualized using metagraph, dashed lines indicate that there are shared nodes. (**II**) Same number of pathways visualized as an interaction network. The size of the node is reduced to improve the readability.

cancers (Figure 2)] as a single metanode and to display the many possible relations between metanodes (Figures 2 and 3) while preserving network topology.

As shown in Figure 3, increased readability and performance are achieved by hiding the unnecessary pathway detail within collapsed metanodes. A global view of the five signaling pathways for *Homo sapiens* becomes very clear when collapsed pathway nodes and correlations between them are automatically generated using one of the default aggregation functions implemented in VisANT. At the same time, detailed relationships can be examined with

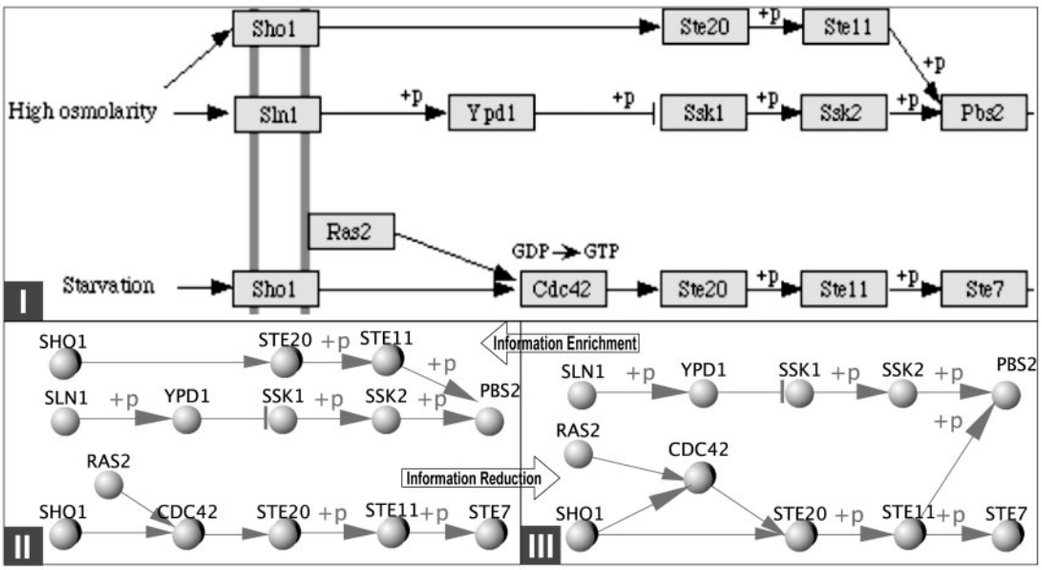

Figure 4: Node duplication preserves information. (**I**) Segment of MAPK signal transduction pathway in yeast visualized using static image in KEGG. There are only one path from SHOI to PBS2: SHOI→STE20→STEII→PBS2. (**II**) Corresponding pathway visualized in VisANT retains the same path through node duplication. (**III**) Visualization of pathway as interaction network lost the information of condition dependency, while results in two different paths from SHOI to PBS2: SHOI→STE20→STEII→PBS2 and SHOI→CDC42→ STE20→STEII→PBS2.

expanded pathways such as the Notch signaling pathway (Figure 3I).

Metagraph related functions are grouped under the ***MetaGraph*** menu in VisANT. Metanodes can be created/removed manually or automatically using an extended edge-list format. Metanodes can also be created based on topological features using built-in functions, for example, disjoint subgraphs can be grouped automatically. The view of the network between different hierarchical levels can be easily carried out using Collapse/Expand. Nodes can be duplicated explicitly and interactions can be easily assigned to a specific instance, which facilitates representation of condition-dependent interactions (e.g. turn Figure 4III into Figure 4II).

It is noteworthy that interaction networks are 'low resolution' networks because they are built by combining pair-wise interactions directly without any specification of their condition dependency or control logic, as opposed to the information inherent in pathways (Figure 4).

FUTURE DEVELOPMENTS

The ultimate goal of VisANT is to be able to provide a visual interface by which large-scale network models that are suitable for modeling and simulation can be built. Achieving this goal will require the incorporation of complete genome annotations and further development of the computational framework which will support such complex models. With this in mind, enhancements to VisANT in the near future will include the following.

Visualization of cellular localization information

Movement between intracellular compartments, or between the interior and exterior of the cell is central to function and to simulating various conditions. Since the set of transitions allowable to a molecule is compartment dependent, the group of entities sharing a single compartment can be represented as metanode, thus avoiding the problem of explicit movement. Hence the same molecule in a cytoplasmic or nuclear metanode will have a different set of available networks.

Visualization of alternative splicing

The considerable amount of alternative splicing now known to occur in higher eukaryotes is one of the surprises of genomics, and its implication in disease processes is particularly striking. Integrating alternative splicing knowledge into VisANT is now becoming practical because this information is now

available in Entrez Gene [48]. The goal is to replace the gene/protein node with a metanode that contains alternative splicing variants. In such a case, genetic interactions will connect the genes in the metanode and protein–protein interactions will connect the splice variants.

Enrichment analyses for pathways or functional modules

If differential expression manifests itself at the level of pathways or functionally coherent sets of genes, it may be more powerful to perform statistical testing for differential expression on a ~~priori~~ a defined gene sets. A representative method is the Gene Set Enrichment Analysis [49–52]. In addition, such analyses will help to determine which pathways/modules are turned on/off for a given expression profile.

Command line mode

This mode allows VisANT to be run without any graphical user interface. VisANT will be able perform specific tasks specified in a command file, allowing users to submit computationally heavy job to super computers. As is evident from its beta release (VisANT 3.19), this mode not only facilitates faster performance and automation, but also increases VisANT's ability to handle large-scale networks, making it possible to process networks with millions of nodes and edges.

Key Points

- VisANT is a software framework for visualizing, mining, analyzing and modeling multi-resolution, multi-scale biological networks.
- VisANT provides gene-based integrative services to automatically resolve ambiguities that result from nonunique protein/gene notation. It is also able to integrate interaction data, splice variants (by using metagraphs), networks and pathways, thereby providing a platform to enable better modeling.
- The system is highly optimized to handle large-scale networks and facilitates the fast loading of large datasets.
- The unique implementation of metagraphs not only enables VisANT to integrate multi-scale knowledge, but also improves network readability and system performance.

Acknowledgements

We thank colleagues, students and other users of VisANT for their valuable comments and suggestions in improving all facets of the system. Figure 2 is prepared with the help of Matt Huyck and Yi-Chien Chang. This work is supported by National Institutes of Health (1R01RR022971-01A1).

References

1. Oltvai ZN, Barabasi AL. Systems biology. life's complexity pyramid. *Science* 2002;**298**:763–4.
2. Kohn KW, Aladjem MI, Weinstein JN, *et al*. Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol Biol Cell* 2006;**17**:1–13.
3. Ravasz E, Somera AL, Mongru DA, *et al*. Hierarchical organization of modularity in metabolic networks. *Science* 2002;**297**:1551–5.
4. Endy D, Brent R. Modelling cellular behaviour. *Nature* 2001;**409**:391–5.
5. Prinz S, Avila-Campillo I, Aldridge C, *et al*. Control of yeast filamentous-form growth by modules in an integrated molecular network. *Genome Res* 2004;**14**:380–90.
6. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA* 2003;**100**:12123–8.
7. Hartwell LH, Hopfield JJ, Leibler S, *et al*. From molecular to modular cell biology. *Nature* 1999;**402**:C47–52.
8. Rives AW, Galitski T. Modular organization of cellular networks. *Proc Natl Acad Sci USA* 2003;**100**:1128–33.
9. Breitkreutz BJ, Stark C, Tyers M. Osprey: a network visualization system. *Genome Biol* 2003;**4**:R22.
10. Junker BH, Klukas C, Schreiber F. VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 2006;**7**:109.
11. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504.
12. Zanzoni A, Montecchi-Palazzi L, Quondam M, *et al*. MINT: a Molecular INTeraction database. *FEBS Lett* 2002;**513**:135–40.
13. Schwarz R, Liang C, Kaleta C, *et al*. Integrated network reconstruction, visualization and analysis using YANAsquare. *BMC Bioinformatics* 2007;**8**:313.
14. Bourqui R, Cottret L, Lacroix V, *et al*. Metabolic network visualization eliminating node redundance and preserving metabolic pathways. *BMC Syst Biol* 2007;**1**:29.
15. Negishi Y, Nakamura H, Yachie N, *et al*. eXpanda: an integrated platform for network analysis and visualization. *In Silico Biol* 2007;**7**:141–4.
16. Suderman M, Hallett M. Tools for visually exploring biological networks. *Bioinformatics* 2007;**23**:2651–9.
17. Baitaluk M, Qian X, Godbole S, *et al*. PathSys: integrating molecular interaction graphs for systems biology. *BMC Bioinformatics* 2006;**7**:55.
18. Iragne F, Nikolski M, Mathieu B, *et al*. ProViz: protein interaction visualization and exploration. *Bioinformatics* 2005;**21**:272–4.
19. Hu Z, Mellor J, DeLisi C. Analyzing networks with VisANT. In: Baxevanis A, Davison D, Page R, *et al*, (eds). *Current Protocols in Bioinformatics*. New Jersey: John Wiley & Sons, Inc., 2004.
20. Hu Z, Mellor J, Wu J, *et al*. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* 2004;**5**:17.
21. Hu Z, Mellor J, Wu J, *et al*. Towards zoomable multidimensional maps of the cell. *Nat Biotechnol* 2007;**25**:547–54.

22. Hu Z, Mellor J, Wu J, *et al*. VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* 2005;**33**:W352–7.
23. Hu Z, Ng DM, Yamada T, *et al*. VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Res* 2007;**35**:W625–32.
24. Hermjakob H, *et al*. The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat Biotechnol* 2004;**22**: 177–83.
25. Cline MS, Smoot M, Cerami E, *et al*. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2007;**2**:2366–82.
26. Mellor JC, Yanai I, Clodfelter KH, *et al*. Predictome: a database of putative functional links between proteins. *Nucleic Acids Res* 2002;**30**:306–9.
27. Goh CS, Bogan AA, Joachimiak M, *et al*. Co-evolution of proteins with their interaction partners. *J Mol Biol* 2000;**299**: 283–93.
28. von Mering C, Jensen LJ, Kuhn M, *et al*. STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* 2007;**35**: D358–62.
29. Kanehisa M, Araki M, Goto S, *et al*. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008;**36**:D480–4.
30. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–9.
31. Mewes HW, Dietmann S, Frishman D, *et al*. MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res* 2008;**36**:D196–201.
32. Bader GD, Betel D, Hogue CW. BIND: the biomolecular interaction network database. *Nucleic Acids Res* 2003;**31**: 248–50.
33. Mishra GR, Suresh M, Kumaran K, *et al*. Human protein reference database—2006 update. *Nucleic Acids Res* 2006;**34**: D411–4.
34. Bleakley K, Biau G, Vert JP. Supervised reconstruction of biological networks with local models. *Bioinformatics* 2007; **23**:i57–65.
35. Jansen R, Yu H, Greenbaum D, *et al*. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 2003;**302**:449–53.
36. Benson DA, Karsch-Mizrachi I, Lipman DJ, *et al*. GenBank. *Nucleic Acids Res* 2008;**36**:D25–30.
37. Hamosh A, Scott AF, Amberger J, *et al*. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2002;**30**:52–5.
38. The universal protein resource (UniProt). *Nucleic Acids Res* 2008;**36**:D190–5.
39. Bruford EA, Lush MJ, Wright MW, *et al*. The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res* 2008;**36**:D445–8.
40. Breitkreutz BJ, Stark C, Reguly T, *et al*. The BioGRID interaction database: 2008 update. *Nucleic Acids Res* 2008;**36**: D637–40.
41. Schuster-Bockler B, Bateman A. Protein interactions in human genetic diseases. *Genome Biol* 2008;**9**:R9.
42. Yeger-Lotem E, Sattath S, Kashtan N, *et al*. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci USA* 2004;**101**:5934–9.
43. Zhang LV, King OD, Wong SL, *et al*. Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. *J Biol* 2005;**4**:6.
44. Cherry JM, Adler C, Ball C, *et al*. SGD: *Saccharomyces genome* database. *Nucleic Acids Res* 1998;**26**:73–9.
45. Wilson RJ, Goodman JL, Strelets VB. FlyBase: integration and improvements to query tools. *Nucleic Acids Res* 2008;**36**: D588–93.
46. Rogers A, Antoshechkin I, Bieri T, *et al*. WormBase 2007. *Nucleic Acids Res* 2008;**36**:D612–7.
47. Goh KI, Cusick ME, Valle D, *et al*. The human disease network. *Proc Natl Acad Sci USA* 2007;**104**:8685–90.
48. Maglott D, Ostell J, Pruitt KD, *et al*. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2005;**33**: D54–58.
49. De Windt A, Rai M, Bernier L, *et al*. Gene set enrichment analysis reveals several globally affected pathways due to SKI-1/S1P inhibition in HepG2 cells. *DNA Cell Biol* 2007; **26**:765–72.
50. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;**102**:15545–50.
51. Lewin A, Grieve IC. Grouping Gene Ontology terms to improve the assessment of gene set enrichment in microarray data. *BMC Bioinformatics* 2006;**7**:426.
52. Klebanov L, Glazko G, Salzman P, *et al*. A multivariate extension of the gene set enrichment analysis. *J Bioinform Comput Biol* 2007;**5**:1139–53.