Article

Network informatics: A new science

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; International Academy of Ecology and Environmental Sciences, Hong Kong
E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 6 August 2015; *Accepted 8 September 2015*; *Published online 1 June 2016*

Abstract
Based on my previous study, in present article I further outlined and defined the aims, scope, theory and methodology of network informatics.

Keywords network informatics; methodology; theory; scientific discipline.

Selforganizology
ISSN 2410-0080
URL: http://www.iaees.org/publications/journals/selforganizology/online-version.asp
RSS: http://www.iaees.org/publications/journals/selforganizology/rss.xml
E-mail: selforganizology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Organisation and communication of information in a network is considerably influenced by network properties, network structure, and network dynamics, etc. As a result, I first proposed the science concept, network informatics (Zhang, 2016c). Network informatics aims to understand to investigate the structure, properties and organization of scientific information in the network view. In present study, I further outlined and defined the aims, scope, theory and methodology of network informatics.

2 Aims and Scope

Network informatics is an interdisciplinary science based on informatics, network science, and other related scientific disciplines. In particular, it is a network-based science, as other new proposed sciences (Zhang, 2016c). Network informatics aims to understand and investigate the structure, properties and organization of information in the network. The scope of network informatics covers but not limits to: (1) theories, algorithms and software of network informatics; (2) mechanisms and rules of flow and organization of information in the network; (3) theory and methodology of dynamics, optimization and control of information networks; (4) network analysis of information networks; (5) factors that affect organization and communication of information, etc.

3 Scientific Foundation

3.1 Informatics

Informatics is the science which investigates the structure and properties of scientific information, the regularities of scientific information activity, its theory, methodology and organization (Mikhailov et al., 1966). Informatics also studies the systems that represent, process, and communicate information. Later, informatics was defined as the study of the structure, algorithms, behaviour, and interactions of natural and artificial computational systems by the University of Edinburgh in 1994 (Wikipedia, 2016a).

Informatics considers the interaction between humans and information alongside the construction of interfaces organisation, technology and system. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields (Wikipedia, 2016a). Informatics covers such areas as computer science, information system, mathematics and statistics, information technology, biology, socialogy, etc.

One of the most significant areas of applied informatics is organizational informatics. Organizational informatics is fundamentally interested in the application of information, information systems and ICT within organisations of various forms including private sector, public sector and voluntary sector organisations (Beynon-Davies, 2002, 2009).

3.2 Network science

Network science is a science which studies complex networks such as telecommunication networks, computer networks, biological networks, cognitive networks, and social networks, considering distinct elements or actors represented by nodes and the connections between the elements or actors as links (Wikipedia, 2016b). Network science draws on theories and methods including graph theory from mathematics, statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, and social structure from sociology (Wikipedia, 2016b). Also, the United States National Research Council defined network science as "the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena." (Committee on Network Science for Future Army Applications, 2006)

4 Methodology

Based on high-throughput -omics data, network database retrievals and other information, network informatics stresses construction of information networks (i.e., biological networks, communication networks, etc.), topological analysis of information networks, network flow analysis, structural optimization and optimal control of information networks, etc (Zhang, 2016g).

4.1 Data source

There are two sources of data for research in network informatics, public databases and experimental verification. First, we can use public databases, i.e., the existing public data and published data, to construct network models of the information network and analyze intrinsic mechanism, and finally validate the mechanism through experiments (Zhou et al., 2012). Second, we may use various technologies to investigate the interactions between the toxicant and network model, to construct and analyze information network based on the generated data, and to analyze the mechanism of information organization and communication.

4.2 Big data analytics

Big data is the data sets so large or complex that conventional data processing tecgniques are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy (Wikipedia, 2016c; Zhang, 2016g).

Big data analytics is the process of examining big data to uncover hidden patterns, unknown correlations and other useful information. With big data analytics, e.g., high-performance data mining, predictive analytics, text mining, forecasting and optimization, we can analyze huge volumes of data that conventional analytics can not handle. In addition, machine learning techaniques are ideally suited to addressing big data needs (Zhang, 2007b, 2010; Zhang and Qi, 2014; SAS, 2016). Many problems in network informatics are expected to be addressed by using big data analytics.

4.3 Network construction and interactions prediction

An information network is the most important basis for further network informatics studies. How to find interactions and construct an information network is necessary. Zhang (2011a, 2012a, 2012b) has proposed a series of correlation methods to construct networks. Pearson correlation measure will lead to a false result (Zhang and Li, 2015). Thus, Zhang (2015c) used partial linear correlation and proposed some partial correlation measures, and used them to jointly predict interactions (Zhang, 2015b). Moreover, there are a lot of other studies on construction and prediction of biological networks (Goh et al., 2000; Pazos and Valencia, 2001; Guimera and Sales-Pardo, 2009).

We may use an incomplete network to predict missing interactions (links) (Clauset et al., 2008; Guimera and Sales-Pardo, 2009; Barzel and Barabási, 2013; Lü et al., 2015; Zhang, 2015d, 2016a, 2016d; Zhang and Li, 2015).

It is expected that network evolution based (Zhang, 2012a, 2015a, 2016b), node similarity based (Zhang, 2015d; based on prediction from primary structure), and correlation based (Zhang, 2007a, 2011a, 2012a, 2012b, 2015d, 2016d; Zhang and Li, 2015) methods are expected to be the most promising in the future.

4.4 Network analysis

Network analysis covers a variety of areas and methods (Zhang, 2012a). Main contents of network analysis, to be used in network informatics, include the following aspects.

4.4.1 Attribute analysis

Attribute analysis aims to screen node attributes (e.g., protein attributes, etc.) based on their contribution to topological structure of the network (Zhang, 2016e).

4.4.2 Topological analysis

Topological analysis of networks mainly includes the following

Find trees in the network: DFS algorithm, Minty’s algorithm, etc (Minty, 1965; Zhang, 2012a).

Find circuits (closed paths, loops) (Paton, 1969; Zhang, 2012a, 2016e).

Finf the maximal flow: Ford—Fulkerson algorithm (Ford and Fulkerson , 1956; Zhang, 2012a).

Find the shortest path: Dijkstra algorithm, Floyd algorithm (Dijkstra, 1959; Zhang, 2012a; Zhang, 2016e).

Find the shortest tree: Kruskal algorithm (Zhang, 2012a).

Calculate network connectedness (connectivity), blocks, cut vertices, and bridges (Zhang, 2012a).

Calculate node centrality (Zhang, 2012a, 2012c; Shams and Khansari, 2014; Jesmin et al., 2016).

Find modules, mosaics, and sub-networks (Kondoh, 2008; Bascompte, 2009; Zhang, 2016f; Zhang and Li, 2016).

Analyze degree distribution (Huang and Zhang, 2012; Zhang, 2011a, 2012a, 2012c; Zhang and Zhan, 2011; Rahman et al., 2013).

For example, degree distribution and crucial metabolites/reactions of tumor pathways have been conducted (Huang and Zhang, 2012; Li and Zhang, 2013; Zhang, 2012c). In addition to the methods above, other statistical methods, e.g., PCA, etc., are also useful in network analysis.

4.4.3 Network structure and stability

Stability of biological networks has been studied in the past (Din, 2014). These studies have been focused on ecosystems and the methods can be used in the phamarceutical studies. Pinnegar et al. (2005) used a detailed Ecopath with Ecosim (EwE) model to test the impacts of food web aggregation and the removal of weak linkages. They found that aggregation of a 41-compartment food web to 27 and 16 compartment systems greatly affected system properties (e.g. connectance, system omnivory, and ascendancy) and influenced dynamic stability (Zhang, 2012a).

The most developed theory is that there is a relationship between network connectance and different types of ecosystem stability. Some models suggest that lower connectance involve higher local (May, 1973; Pimm, 1991; Chen and Cohen, 2001) and global (Cohen et al., 1990; Chen and Cohen, 2001) stability, i.e., the system recovers faster after a disturbance. However, another theory suggests that a food web with higher connectance has more numerous reassembly pathways and can thus recover faster from perturbation (Law and Blackford, 1992).

4.4.4 Flow (flux) balance analysis

Flow balance analysis aims to analyze network flows at steady state. Differential equations and other equations are usually used to describe network dynamics (Chen et al., 2010; Schellenberger et al., 2011). As an example, Jain et al. (2011) used mathematical models to decipher balance between cell survival and cell death using insulin.

Some standardized indices and matrices can be used in flow balance analysis (Latham, 2006; Fath et al., 2007; Zhang, 2012a). They include Average Mutual Information (AMI) (Rutledge et al., 1976). Ascendency (A) index of a system was developed by Ulanowicz (1983, 1997). Compartmentalization index is used to measure the degree of well-connected subsystems within a network (Pimm and Lawton, 1980). Constraint efficiency is a measure of a total of constraints that govern flow out of individual compartments (Latham and Scully, 2002). Zorach and Ulanowicz (2003) have presented effective measures (effective connectivity, effective flows, effective nodes, effective rules) for weighted networks. Fath and Patten (1999a) developed a measure (measures the evenness of flow in a network) for network homogenization. In addition, Higashi and Patten (1986, 1989) and Fath and Patten (1999b) presented an index for describing the dominance of indirect effects.

4.4.5 Network models

Some network models have been developed for food webs (Zhang, 2012a), such as cascade model (Cohen et al., 1990), niche model (Williams and Martinez, 2000), multitrophicassembly model (Pimm 1980, Lockwood et al. 1997), MaxEnt models (Williams, 2010), and Ecopath model (Polovina, 1984; Christensen and Pauly, 1992; Libralato et al., 2006), etc. Ecosim is the dynamic program of the EwE (Walters et al., 1997, 2000). It is based on a set of differential equations derived from the Ecopath equation above, which allows a dynamic representationof the system variables, like biomasses, predation, and production (Libralato et al., 2006). They can be revised and improved to fit information networks.

4.5 Network dynamics, evolution and control

Ferrarini (2011a, 2011b, 2013a-d, 2014) have proposed a series of thoughts and methods on the dynamics, controllability and dynamic control of biological networks. Zhang (2015a) proposed a generalized network evolution model and self-organization theory on community assembly, in which the model is a series of differential (difference) equations with different number as the time. In addition, Zhang (2016b) developed a random network based, node attraction facilitated network evolution method. The two dynamic models are useful to study the network evolution, dynamics, and to predict interactions.

Network is optimized to search for an optimal search plan, and achieve a topological structure so that the network possesses relative stability (Zhang, 2012a).

The dynamic control of network means to change topological structure and key parameters of the network stage by stage so that the goal function of entire network achieves the optimum or suboptimum (Zhang, 2012a). Mathematical tools, like dynamic programming, decision-making analysis, game theory, etc., can be used to address these problems.

4.6 Network visualization

Network visualization aims to present users with the static/dynamic two- or three-dimensional illustrations and images of information networks. There are a variety of such network software for doing it (Zhang, 2012a), for example, ABNNSim (Schoenharl, 2005), Topographica (Bednar et al., 2004), Pajek, NetDraw, NetLogo (Resnick, 1994), netGenerator (Zhang, 2012a, 2012d), Repast (Macal and North, 2005), Topographica (Bednar et al., 2004), Startlogo (Resnick, 1994), etc.

Acknowledgment

We are thankful to the support of Discovery and Crucial Node Analysis of Important Biological and Social Networks (2015.6-2020.6), from Yangling Institute of Modern Agricultural Standardization, and High-Quality Textbook *Network Biology* Project for Engineering of Teaching Quality and Teaching Reform of Undergraduate Universities of Guangdong Province (2015.6-2018.6), from Department of Education of Guangdong Province.

References

Barzel B, Barabási AL. 2013. Network link prediction by global silencing of indirect correlations. Nature Biotechnology, 31: 720-725

Bednar JA, Choe Y, Paula JD, et al. 2004. Modeling cortical maps with topographica. Neurocomputing, 58: 1129-1135

Beynon-Davies P. 2002. Information Systems: An Introduction to Informatics in Organisations. Palgrave, Basingstoke, UK

Beynon-Davies P. 2009. Business Information Systems. Palgrave, Basingstoke, UK

Chen Q, Wang Z, Wei DQ. 2010. Progress in the applications of flux analysis of metabolic networks. Chinese Science Bulletin, 2010, 55(22): 2315-2322

Chen X, Cohen JE. 2001. Global stability, local stability and permanence in model food webs. Journal of Theoretical Biology, 212: 223-235

Clauset A, Moore C, Newman MEJ. 2008. Hierarchical structure and the prediction of missing links in networks. Nature, 453: 98-101

Committee on Network Science for Future Army Applications. 2006. Network Science. National Research Council

Dijkstra EW. 1959. A note on two problems in connexion with graphs. Numerischemathematik, 1(1): 269-271

Din Q. 2014. Stability analysis of a biological network. Network Biology, 4(3): 123-129

Fath BD, Patten BC. 1999a. Quantifying resource homogenization using network flow analysis. Ecological Modelling, 123: 193-205

Fath BD, Patten BC. 1999b. Review of the foundations of network environ analysis. Ecosystems, 2: 167-179

Fath BD, Scharler UM, Ulanowiczd RE, Hannone B. 2007. Ecological network analysis: network construction. Ecological Modeling, 208: 49-55

Ferrarini A. 2011a. Some thoughts on the controllability of network systems. Network Biology, 1(3-4): 186-188

Ferrarini A. 2011b. Some steps forward in semi-quantitative network modelling. Network Biology, 1(1): 72-78

Ferrarini A. 2013a. Exogenous control of biological and ecological systems through evolutionary modelling. Proceedings of the International Academy of Ecology and Environmental Sciences, 3(3): 257-265

Ferrarini A. 2013b. Controlling ecological and biological networks via evolutionary modelling. Network Biology, 3(3): 97-105

Ferrarini A. 2013c. Computing the uncertainty associated with the control of ecological and biological systems. Computational Ecology and Software, 3(3): 74-80

Ferrarini A. 2013d. Networks control: introducing the degree of success and feasibility. Network Biology, 3(4): 115-120

Ferrarini A. 2014. Local and global control of ecological and biological networks. Network Biology, 4(1): 21-30

Ford LR Jr, Fulkerson DR. l956. Maxima1 flow through a network. Canadian Journal of Mathematics, 8: 399-404

Guimera R, Sales-Pardo M. 2009. Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences of USA, 106: 22073-22078

Huang JQ, Zhang WJ. 2012. Analysis on degree distribution of tumor signaling networks. Network Biology, 2(3): 95-109

Jain S, Bhooshan SV, Naik PK. 2011. Mathematical modeling deciphering balance between cell survival and cell death using insulin. Network Biology, 1(1): 46-58

Jiang LQ, Zhang WJ. 2015. Determination of keystone species in CSM food web: A topological analysis of network structure. Network Biology, 5(1): 13-33

Latham LG, Scully EP. 2002. Quantifying constraint to assess development in ecological networks. Ecological Modelling, 154: 25-44

Latham LG. 2006. Network flow analysis algorithms. Ecological Modelling, 192: 586-600

Law R, Blackford JC. 1992. Self-assembling food webs. A global view-point of coexistence of species in Lotka-Volterra communities. Ecology, 73: 567–578

Li JR, Zhang WJ. 2013. Identification of crucial metabolites/reactions in tumor signaling networks. Network Biology, 3(4): 121-132

Lü LY, Pan LM, Zhou T, et al. 2015. Toward link predictability of complex networks. Proceedings of the National Academy of Sciences of USA, 112: 2325-2330

Macal CM, North MJ. 2005. Tutorial on agent-based modeling and simulation. In: Proceedings of the 2005 Winter Simulation Conference (Kuhl ME, Steiger NM, Armstrong FB, Joines JA, eds).

May RM. 1973. Stability and complexity in model ecosystems. Princeton University Press, USA

Mikhailov AI, Chernyl AI, Gilyarevskii RS. 1966. Informatika – novoe nazvanie teorii naučnoj informacii. Naučno tehničeskaja informacija, 12: 35-39

Minty GJ. 1965. A simple algorithm for listing all the trees of a graph. IEEE Trans on Circuit Theory, CT-12(1): 120

Paton K. 1969. An algorithm for finding a fundamental set of cycles of a graph. Communications of the ACM, 12(9): 514-518

Pimm SL. 1991. The Balance of Nature. University of Chicago Press, Chicago, USA

Pinnegar JK, Blanchard JL, Mackinson S, et al. 2005. Aggregation and removal of weak-links in food-web models: system stability and recoveryfrom disturbance. Ecological Modelling, 184: 229-248

Resnick M. 1994. Turtles, Termites and Traffic Jams. MIT Press, USA

Rutledge RW, Basorre BL, Mulholland RJ. 1976. Ecological stability: an information theory viewpoint. Journal of Theoretical Biology, 57: 355-371

Schoenharl TW. 2005. An Agent Based Modeling Approach for the Exploration of Self-organizing Neural Networks. MS Thesis, University of Notre Dame, USA

Shams B, Khansari M. 2014. Using network properties to evaluate targeted immunization algorithms. Network Biology, 4(3): 74-94

Ulanowicz RE. 1983. Identifying the structure of cycling in ecosystems. Mathematical Biosciences, 65: 219-237

Ulanowicz RE. 1997. Ecology, the ascendent perspective. In: Complexity in Ecological Systems Series (Allen TFH, Roberts DW, eds). Columbia University Press, New York, USA

Williams RJ. 2010. Simple MaxEnt models explain food web degree distributions. Theoretical Ecology, 3: 45-52

Wikipedia. 2016a. Informatics. https://en.wikipedia.org/wiki/Informatics. Accessed February 3, 2016

Wikipedia. 2016b. Network science. https://en.wikipedia.org/wiki/Network_science. Accessed February 3, 2016

Zhang WJ. 2007. Computer inference of network of ecological interactions from sampling data. Environmental Monitoring and Assessment, 124: 253-261

Zhang WJ. 2011a. Constructing ecological interaction networks by correlation analysis: hints from community sampling. Network Biology, 1(2): 81-98

Zhang WJ. 2011b. Network Biology: an exciting frontier science. Network Biology, 1(1): 79-80

Zhang WJ. 2012a. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2012b. How to construct the statistic network? An association network of herbaceous plants constructed from field sampling. Network Biology, 2(2): 57-68

Zhang WJ. 2012c. Several mathematical methods for identifying crucial nodes in networks. Network Biology, 2(4): 121-126

Zhang WJ. 2012d. A Java software for drawing graphs. Network Biology, 2(1): 38-44

Zhang WJ. 2013. Selforganizology: A science that deals with self-organization. Network Biology, 3(1):1-14

Zhang WJ. 2014. Selforganizology: A more detailed description. Selforganizology, 1(1): 31-46

Zhang WJ. 2015a. A generalized network evolution model and self-organization theory on community assembly. Selforganizology, 2(3): 55-64

Zhang WJ. 2015b. A hierarchical method for finding interactions: Jointly using linear correlation and rank correlation analysis. Network Biology, 5(4): 137-145

Zhang WJ. 2015c. Calculation and statistic test of partial correlation of general correlation measures. Selforganizology, 2(4): 65-77

Zhang WJ. 2015d. Prediction of missing connections in the network: A node-similarity based algorithm. Selforganizology, 2(4): 91-101

Zhang WJ. 2016a. A node degree dependent random perturbation method for prediction of missing links in the network. Network Biology, 2016, 6(1): 1-11

Zhang WJ. 2016b. A random network based, node attraction facilitated network evolution method. Selforganizology, 3(1): 1-9

Zhang WJ. 2016c. Network chemistry, network toxicology, network informatics, and network behavioristics: A scientific outline. Network Biology, 6(1): 37-39

Zhang WJ. 2016d. Selforganizology: The Science of Self-Organization. World Scientific, Singapore

Zhang WJ. 2016e. Screening node attributes that significantly influence node centrality in the network. Selforganizology, 3(3)

Zhang WJ. 2016f. A method for identifying hierarchical sub-networks and weighting network links based on their similarity in sub-network affiliation. Selforganizology, 3(2): 43-53

Zhang WJ. 2016g. Network toxicology: A new science. Network Biology, 6(2): 40-49

Zhang WJ, Li X. 2015. Linear correlation analysis in finding interactions: Half of predicted interactions are undeterministic and one-third of candidate direct interactions are missed. Selforganizology, 2(3): 39-45

Zhang WJ, Li X. 2016. A cluster method for finding node sets / sub-networks based on between- node similarity in sets of adjacency nodes: with application in finding sub-networks in tumor pathways. Proceedings of the International Academy of Ecology and Environmental Sciences, 6(1): 13-23

Zhang WJ, Zhan CY. 2011. An algorithm for calculation of degree distribution and detection of network type: with application in food webs. Network Biology, 1(3-4): 159-170

Zhou WX, Cheng XR, Zhang YX. 2012. Network pharmacology-a new philosophy for understanding of drug action and discovery of new drugs. Chinese Journal of Pharmacology and Toxicology, 2(26): 4-9