

# Bioinformatics Center – Mathematical Bioinformatics –

<http://www.bic.kyoto-u.ac.jp/takutsu/>



Prof  
AKUTSU, Tatsuya  
(D Eng)



Assist Prof  
HAYASHIDA, Morihiro  
(D Inf)



Assist Prof  
TAMURA, Takeyuki  
(D Inf)

## Students

NAKAJIMA, Natsu (D3)

ZHAO, Yang (D3)

LU, Wei (D3)

HASEGAWA, Takanori (D2)

MORI, Tomoya (D2)

RUAN, Peiyong (D2)

UECHI, Risa (D2)

JIRA, Jindalertudomdee (D1)

CONG, Xiao (M2)

HWANG, Jaewook (M2)

BAO, Yu (RS)

CAO, Yue (RS)

## Visiting Researchers

Ms. QIU, Yushan

Ms. TASSINARI, Anna

Mr. CHANG, Chia-Jung

Assoc Prof CAI, Hongmin

Senior Researcher VERT, Jean-Philippe

Assoc Prof Em MELKMAN, Avraham Aharon

The University of Hong Kong, China, P.R., 20 May–15 August

Boston University, U.S.A. 27 June–23 September

National Taiwan University, Taiwan, 10 July–3 September

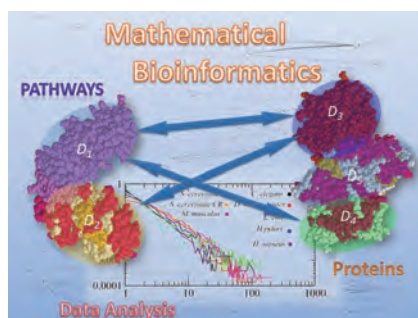
South China University of Technology, China, P.R., 30 May–23 August

Curie Institute, France, 27 June–25 August

Ben Gurion University of the Negev, Israel, 20 February–24 April

## Scope of Research

Due to rapid progress of the genome projects, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are recently studying the following topics: systems biology, scale-free networks, protein structure prediction, inference of biological networks, chemo-informatics, discrete and stochastic methods for bioinformatics.



## KEYWORDS

Scale-free Networks  
Boolean Networks  
Grammar-based Compression  
RNA Secondary Structures  
Chemical Graphs



## Selected Publications

Akutsu, T.; Fukagawa, D.; Halldorsson, M. M.; Takasu, A.; Tanaka, K., Approximation and Parameterized Algorithms for Common Subtrees and Edit Distance between Unordered Trees, *Theoretical Computer Science*, **470**, 10-22 (2013).

Akutsu, T.; Tamura, T., A Polynomial-time Algorithm for Computing the Maximum Common Connected Edge Subgraph of Outerplanar Graphs of Bounded Degree, *Algorithms*, **6**, 119-135 (2013).

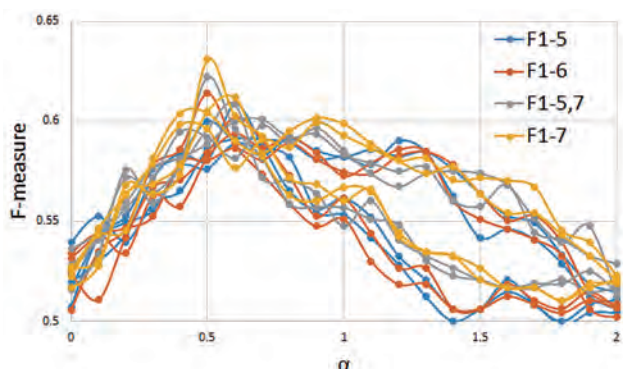
Nacher, J. C.; Akutsu, T., Structural Controllability of Unidirectional Bipartite Networks, *Scientific Reports*, **3**, 1647 (2013).

Zhao, Y.; Tamura, T.; Akutsu, T.; Vert, J.-P., Flux Balance Impact Degree: A New Definition of Impact Degree to Properly Treat Reversible Reactions in Metabolic Networks, *Bioinformatics*, **29**, 2178-2185 (2013).

Ruan, P.; Hayashida, M.; Maruyama, O.; Akutsu, T., Prediction of Heterodimeric Protein Complexes from Weighted Protein-Protein Interaction Networks Using Novel Features and Kernel Functions, *PLoS ONE*, **8**, e65265 (2013).

## Prediction of Heterodimeric Protein Complexes from Weighted Protein-Protein Interaction Networks Using Novel Features and Kernel Functions

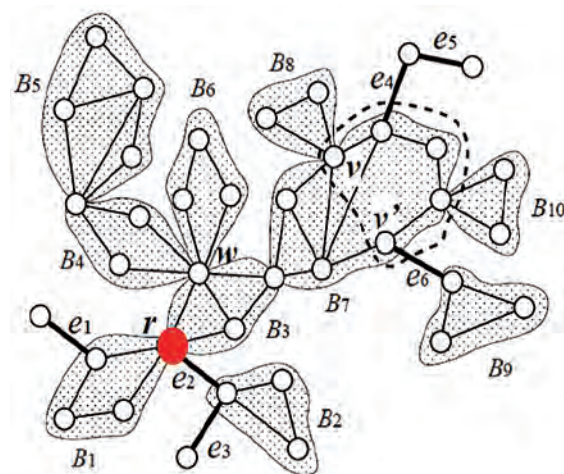
Identification of protein complexes is very important because many proteins express their functional activity by interacting with other proteins and forming protein complexes. Hence, many prediction methods for protein complexes from protein-protein interactions have been developed such as MCL, MCODE, RNSC, PCP, RRW, and NWE. These methods have dealt with only complexes with size of more than three because the methods often were developed based on some density of subgraphs. Heterodimeric protein complexes that consist of two distinct proteins, however, occupy a large part of whole protein complexes according to several comprehensive databases of known complexes. In this study, we proposed seven feature space mappings (F1),..., (F7) from protein-protein interaction data, in which each interaction is weighted based on reliability. Furthermore, we made use of prior knowledge on protein domains, and proposed domain composition kernel  $K_c(C_i, C_j)$  for sets  $C_i, C_j$  of two distinct proteins and its combination kernel  $\langle \varphi(C_i), \varphi(C_j) \rangle + \alpha K_c(C_i, C_j)$  with our proposed features  $\varphi$ . Here,  $K_c(C_i, C_j) = 1$  if the domain composition of  $C_i$  is equivalent to that of  $C_j$ , otherwise 0. We performed ten-fold cross-validation computational experiments for WI-PHI protein-protein interaction data and CYC2008 protein complex catalog, and calculated the F-measures. These results suggest that our proposed kernel considerably outperforms the naive Bayes-based method, which is the best existing method for predicting heterodimeric protein complexes.



**Figure 1.** Result on the average F-measure using four sets (F1-5), (F1-6), (F1-5,7), (F1-7) of features and the domain composition kernel with combination parameter  $\alpha=0, \dots, 2$ .

## Maximum Common Connected Edge Subgraph Problem for Chemical Compounds

Calculating similarity of two given chemical compounds is important and fundamental task in chemical informatics. The maximum common connected edge subgraph problem (MCCES) is to find a connected graph with the maximum number of edges that is isomorphic to a subgraph of each of the two input graphs. Though MCCES is useful for measuring similarity of chemical compounds, MCCES is NP-hard even for labeled partial k-trees of bounded degree, and it is reported that most chemical compounds have treewidth at most 3. On the other hand, it is also reported that 94.4% of chemical compounds have outerplanar graph structures, and the maximum degree of almost all chemical compounds is bounded by a constant (e.g., 8). In this study, we developed a polynomial time algorithm for MCCES when the two input graphs are outerplanar graphs of a bounded vertex degree, where it is known that the problem is NP-hard, even for outerplanar graphs of an unbounded degree.



**Figure 2.** A graph is outerplanar if it can be drawn on a plane such that all vertices lie on the outer face without crossing of edges. Our developed algorithm treats outerplanar graphs as trees, and then dynamic programming-based methods are applied.