

## 最小支配集合に基づく有向生体ネットワーク解析のための高速アルゴリズム

石塚雅之<sup>1</sup> 阿久津達也<sup>2</sup> ナチエル・ホセ<sup>3</sup>

複雑ネットワーク研究において、制御理論と複雑ネットワークを組み合わせることの重要性が認識され、最大マッチング(MM)及び最小支配集合(MDS)などを用いたドライバー頂点の同定法が開発されてきた。近年の研究の進展によりプロテオームワイドなタンパク質間相互作用ネットワークを始めとする大規模無向ネットワークに対するMDSの計算が可能となったが、遺伝子制御ネットワークや代謝経路、シグナル伝達経路など、生体ネットワークの多くは有向である。そこで本研究では、有向生体ネットワークにおける可制御性解析とそれに基づく頂点の分類を高速に実行するために、厳密性を損なわずにネットワークサイズを大幅に縮小させるヒューリスティクスを組み込んだ新たなアルゴリズムを開発し、その有効性を評価した。

### A fast algorithm for minimum dominating set-based analysis of directed biological networks

MASAYUKI ISHITSUKA<sup>1</sup> TATSUYA AKUTSU<sup>2</sup> JOSE C. NACHER<sup>3</sup>

Network controllability research has shed light on the importance of combining control theory concepts with complex networks to identify a driver set of nodes using maximum matching (MM) and minimum dominating set (MDS) approaches. In particular, recent developments have made it possible to apply MDS approach to large-scale undirected networks such as proteome-wide protein interaction network. However, many real-world biological networks are directed such as gene regulatory networks, metabolic pathways and signal pathways. To efficiently examine controllability features and control categories in large directed biological networks, we have developed a new algorithm that includes heuristics to speed up computation as well as to scale up computable network size.

#### 1. はじめに

最近の研究により、複雑ネットワークにおける可制御性を扱うための新たな理論の設計に有効的な手法が提案された[1][2]。可制御性のモデル化を可能とするフレームワークを有することで、経路やネットワークにおいて制御に関する生物学的プロセスがどのように起こり得るかを、最終的に理解することが可能となる。加えて、特定の生物学的機能やヒトの疾患、そしてネットワーク上の細胞部位に関係した分子集合の同定に貢献することが期待される[3][4][5]。

既存の研究では、最小支配集合(MDS)の手法を用いて複雑ネットワークと制御理論を組み合わせることの有効性が示されている[2][6]。MDSには複数の解が存在するが、全てのMDSに属するノードをCritical、一部のMDSにのみ属するノードをIntermittent、どのMDSにも属さないノードをRedundantにそれぞれ分類することで一意の解を得ることができる[7]。ところが、これら制御カテゴリに分類する既存手法は、MDS問題をノード数と等しい回数解く必要があり、計算量が膨大になってしまう問題がある。無向ネットワークについては、この問題を解決するための新たなアルゴリズムが既に提案されており、トランスクリプトーム情報が統合されたプロテオーム・ワイドなタンパク質間

相互作用ネットワークに適用することで、その可制御性に関する研究が行われている[4]。

しかしながら、現実のネットワークの多くは代謝経路やシグナル経路のような有向ネットワークである。これらのシステムにおける可制御性について効率的な解析を行うためには、計算速度と解析可能ネットワークサイズを向上させるヒューリスティクスを含んだ新たなアルゴリズムが必要である。そこで我々は、MDS手法を用いて有向ネットワークにおける各ノードを制御カテゴリに分類するための新たなアルゴリズムを提案する。このアルゴリズムには上記のヒューリスティクスを含んだ前処理が組み込まれており、計算速度及び解析可能ネットワークサイズを大きく改善している。

#### 2. 研究方法

初めに、有向ネットワークの可制御性における重要な命題を記す[7]。

- 命題2.3 ノード  $v$  が他のノードからリンクされていないならば、 $v$  はCriticalノードである。
- 命題2.4 ノード  $v$  が  $k^{out} = 0$  かつ  $k^{in} = 1$  であるノードを2つ以上リンクしているならば、 $v$  はCriticalノードである。

1 東邦大学大学院 理学研究科 情報科学専攻  
Graduate School of Science, Toho University  
2 京都大学 化学研究所 バイオインフォマティクスセンター  
Bioinformatics Center, Institute for Chemical Research, Kyoto University

3 東邦大学 理学部 情報科学科  
Department of Information Science, Faculty of Science, Toho University

これらの命題は、Critical ノードを同定する前処理にて用いる。また、Redundant ノードの多くは本研究会で示す新たな理論を用いた前処理によって、MDS の計算を行うことなく同定することが可能である。

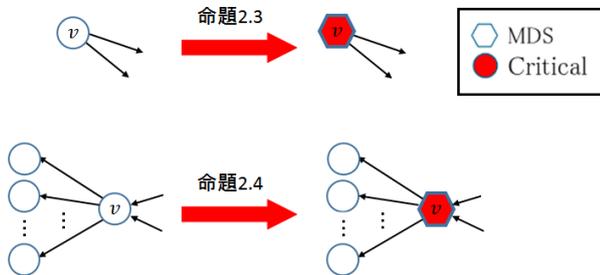


図 1 命題 2.3 と 命題 2.4 を用いた前処理

Figure 1 Pre-processing using proposition 2.3 and proposition 2.4

以上の命題を組み合わせることによって、大規模な有向ネットワークにおける各ノードを効率的に制御カテゴリに分類可能な新たなアルゴリズムを導出した。

- 1 前処理によって得られた情報を整数計画法による定式化に追加し、簡略化された MDS 問題を解く。なお、 $V$  はノードの集合を、 $E$  は有向リンクの集合を示す。

$$\begin{aligned} & \text{minimize} && \sum_{v \in V} x_v \\ & \text{subject to} && x_v + \sum_{(u,v) \in E} x_u \geq 1 \quad \forall v \in V, \quad (\text{Eq. 1}) \\ & && x_v \in \{0,1\} \quad \forall v \in V. \end{aligned}$$

Eq. 1 中の  $(u,v) \in E$  は、ノード  $u$  からノード  $v$  への有向リンクの集合を示している。

- 2 MDS の解を得た後、前処理によって制御カテゴリに分類されなかったノードに対して以下のように制御カテゴリ分類アルゴリズムを適用する。
  - 2.1 Critical ノードに分類するアルゴリズム[7]を、MDS に属するノードに適用する。
  - 2.2 Redundant ノードに分類するアルゴリズム[7]を、MDS に属さないノードに適用する。

### 3. 結果

アルゴリズムの性能を評価するために 500~25,000 ノードから成るスケールフリー・ネットワークを人工的に構築し、計算を行った。その結果、計算可能なネットワークサイズが 25,000 ノードに拡張され、計算速度も  $N=7000$  のネットワークにおいて 156 倍速くなっていることが確認された。

ノード数	計算時間(ms)	
	既存手法	提案手法
1000	$1.24 \times 10^5$	$5.66 \times 10^3$
3000	$1.44 \times 10^5$	$1.79 \times 10^4$
5000	$6.21 \times 10^6$	$4.16 \times 10^4$
7000	$1.16 \times 10^7$	$7.39 \times 10^4$
10000	-	$1.45 \times 10^5$
25000	-	$1.21 \times 10^6$

表 1 ノード数ごとの計算時間

Table 1 Computational time versus network size

### 4. まとめ

アルゴリズムの顕著な性能改善によって、代謝ネットワーク及びシグナル経路のような大規模な有向生体ネットワークにおける MDS 手法を用いた制御性の研究が、初めて可能となった。現在行っている実際の有向ネットワークを用いたデータ解析は、生命システムにおいて重要な制御カテゴリに関連した生物学的機能に対する新たな知見を提供することが期待できる。

### 参考文献

- [1] Liu, Y.-Y.; Slotine, J.-J. & Barabási, A. -L.. Controllability of complex networks. Nature. 2011, vol. 473, p. 167-173.
- [2] Nacher, J. C. and Akutsu, T.. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. New Journal of Physics. 2012, vol. 14, 24 p.
- [3] Wuchty, S.. Controllability in protein interaction networks. Proc. Natl. Acad. Sci. 2014, vol. 111, p. 7156-7160.
- [4] Ishitsuka, M.; Akutsu, T. and Nacher, J. C.. Critical controllability in proteome-wide protein interaction network integrating transcriptome. Scientific Reports. 2016, vol. 6, 23541.
- [5] Nacher, J. C. and Akutsu, T.. Minimum dominating set-based methods for analyzing biological networks. Methods, 2016, vol. 102, p. 57-63.
- [6] Molnár, F.; Sreenivasan, S.; Szymanski, B. K. and Korniss, G.. Minimum Dominating Sets in Scale-Free Network Ensembles. Scientific Reports. 2013, vol. 3, 01736.
- [7] Nacher, J. C. and Akutsu, T.. Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. Journal of Complex Networks. 2014, vol. 2, p. 394-412.