

COMPLEX NETWORKS FROM DUAL POINT OF VIEW

Taichi Haruna^{1,2}

¹Graduate School of Science, Kobe University, 1-1 Rokkodaicho, Nada, Kobe 657-8501, Japan

²PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan

Fax: +81-78-803-5739

tharuna@penguin.kobe-u.ac.jp

(Received October 27, 2010; Accepted November 1, 2010)

(Abstract)

One slogan of science of complex networks is “functionality from network topology”. In order to achieve this slogan, many characteristics such as degree distributions, average path length, clustering coefficients, network motifs etc have been introduced. However, these are all based on the “real view” on networks, that is, nodes are just points and an arc or an edge between two nodes indicates existence of some interaction between the two nodes and nothing more. In this paper we introduce “dual views” on networks in contrast to the “real view” in order to promote further achievement of the slogan. In a “dual view” we interpret objects as processes. This is an abstraction from real systems in which each component has its own process. There are infinitely many “dual views”, however, we can show that there is a canonical “dual view” among all possible ones. We explain mathematical idea to formulate “dual views” and discuss its application to the study of complex networks.

(Keywords)

Complex Networks, Duality, Category Theory, Percolation, Clusterings Comparison

双対的視点による複雑ネットワークへのアプローチ

春名 太一^{1,2}

¹ 神戸大学大学院理学研究科, 〒657-8501
神戸市灘区六甲台町 1-1

² 科学技術振興機構 さきがけ, 〒332-0012
埼玉県川口市本町 4-1-8
tharuna@penguin.kobe-u.ac.jp

1. はじめに

複雑ネットワーク研究の目標の一つはネットワークのトポロジーからその機能を明らかにすることだと考えられる[4]. この目標を達成するためにこれまでに次数分布, 平均頂点間距離, クラスター係数, ネットワークモチーフなど様々な特徴付けの方法が提案されてきた[2-4,16]. ところが, これらの特徴量はいずれも筆者が現実的と呼ぶネットワークの見方に基づいている: つまり, あるシステムがネットワークとして表現されてしまえば, 頂点はただの点でしかなく, 頂点間の

辺もしくは矢印は何らかの相互作用の存在を表す, というごく当たり前の見方である. しかし, 現実のシステムでは頂点として表現される対象内部では何らかのプロセスが走っていることが多い(例えば, 神経ネットワーク, 遺伝子転写制御ネットワークなど). つまり, 「モノをプロセスとして読む」ことができる. このような場合には頂点間の相互作用は内部プロセス同士の接合面であるという双対的な見方ができる[9, 10]. 本稿ではこのような双対的な見方の数学的定式化およびその応用を紹介する. 特に, 双対的な見方に基づいて新たなネットワークの連結性概念を提示し, ネットワークの数理モデルや現実のネットワークの解析への応用について述べる.

「モノをプロセスとして読む」と一口にいってもその読み方は無数にある. また, 双対の取り方として逆の「プロセスをモノとして読む」というやり方もある. 後者の考え方は古くは R. Rosen の Metabolism-Repair System の定式化にも表れているし[20], 最近では複雑ネットワークのコミュニティ同定問題においても使われている[1, 7]. 実は, 両者は圏論的随伴関係と呼ばれる数学的双対性の条件を満たしているのだが[8, 12], 本稿では前者の「モノをプロセスとして読む」方にのみ焦点を絞り, 両者の関係についてはこれ以上深入りしない.

さて, 「モノをプロセスとして読む」やり方は無数にある訳であるが, 少なくとも有向グラフとして表現できるネットワークに関しては, 圏論[14, 15]という数学の枠組みの中で定式化するとすべての双対的視点の中で標準的と呼べるような特定の双対的視点が存在することが示せる[11]. この標準的な双対的視点から有向グラフの矢印の集合上にある同値関係が引き起こされる. この同値関係はすべての双対的視点を考えたときに引き起こされる同値関係の中で最小になっており, ここではこの事実を指して標準的の意味付けを与えているとして差し支えない. 矢印の集合上の同値関係はネットワークの連結性を定義していると考えられ, 標準的な双対的視点から引き起こされる連結性概念を側方連結性と呼ぶ. 本稿では, 側方連結性の導出について圏論を用いずに主にアイデアに焦点を絞り解説する. さらに応用としてネットワークのコンフィギュレーションモデル(頂点の次数分布のみが指定されたラン

ダムネットワーク)上でのパーコレーションを考えることで従来の強連結性・弱連結性との比較を行い、側方連結性の意義を議論する。また、線虫 *C. elegans* の神経ネットワークのデータ解析への応用に関しても議論する。

2. 側方連結性: 有向グラフへの双対的視点から導出される標準的構造

「モノをプロセスとして読む」双対的視点ではネットワークの頂点は内部プロセスとして解釈され、頂点間の相互作用は内部プロセス同士の接合面として理解される。このアイデアは以下のような仕方でもネットワークの変換として数学的に定式化できる。

「1. はじめに」において「モノをプロセスとして読む」方法は無数にあると述べたが、まずは以下のもっとも単純な双対的視点で考えよう。すなわち、ネットワークの頂点を矢印として解釈し、頂点間の相互作用(矢印)は、二つの矢印を結合する頂点として解釈する(Fig.1)。結果として二つの頂点とその間を結ぶ矢印は、この双対的視点の下で、三つの頂点とそれらを直列に繋ぐ二本の矢印へと変換されることになる。実はこの双対的視点で「1. はじめに」で述べた標準的な双対的視点を与えることになる。この双対的視点に伴ってネットワークの変換 L が定義される。 L は頂点を矢印へと変換し、変換前のネットワークにおける矢印が変換後の矢印を繋ぐ糊の役割を果たす(Fig. 2)。

変換 L の形式的な定義は以下になる。まず、有向グラフとは四つ組 $G = (A, O, \partial_0, \partial_1)$ であり、 A は矢印の集合、 O は頂点の集合、 ∂_0, ∂_1 はともに A から O への写像であり、それぞれ与えられた矢印の始点、終点を指定する。与えられた有向グラフ $G = (A, O, \partial_0, \partial_1)$ に対して L による変換後の有向グラフ $L(G)$ は四つ組 $L(G) = (O, O \times \{0,1\} / \sim, \partial'_0, \partial'_1)$ で定義される。ここで、 $\sim$ は集合 $O \times \{0,1\}$ 上の同値関係であり、 $(x,1) \sim (y,0) \Leftrightarrow x$ から y へと向かう矢印が G 内に

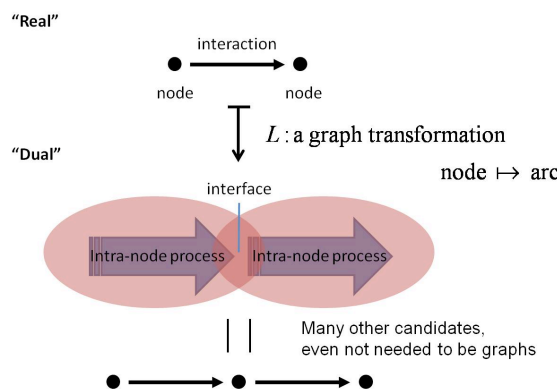


Fig. 1. Switch from the “real view” to a “dual view”.

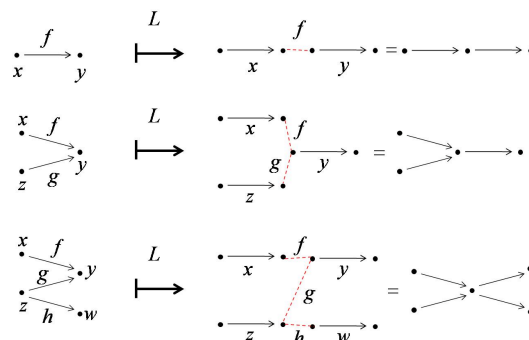


Fig. 2. Some examples for calculation of network transformation L .

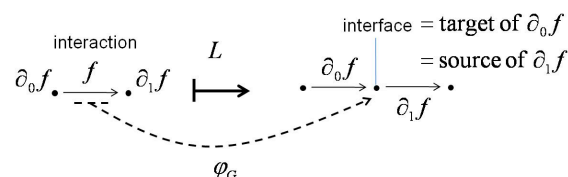


Fig. 3. $\phi_G(f)$ is both target of $\partial_0 f$ and source of $\partial_1 f$.

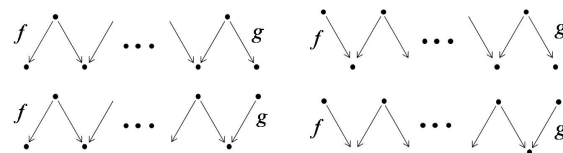


Fig. 4. Arcs f and g are called laterally connected if they are connected by a zigzag sequence of arcs. All four possibilities are indicated.

存在する、という関係により生成される。また、 $\partial_i'(x) = [(x,i)]$ ($i = 0,1$) であり、 $[(x,i)]$ は (x,i) を含む同値類を表す。

さて、変換 L により矢印はその両端の頂点が送られた先の二つの矢印の間を繋ぐ頂点へと送られる。この写像を $\phi_G : A \rightarrow O \times \{0,1\} / \sim$ と書こう。 $f \in A$ に対して $\phi_G(f)$ は $\partial_0 f$ の終点(もしくは $\partial_1 f$ の始点)となっている (Fig.3) ので、 $\phi_G(f) = [(\partial_0 f, 1)] (= [(\partial_1 f, 0)])$ と書ける。写像 ϕ_G に関する自然な問いとして「いつ $\phi_G(f) = \phi_G(g)$ となるか」が考えられる。この問いには容易に答えることができ、矢印 $f, g \in A$ の間に Fig.4 の四つの場合のどれかに対応する矢印の「ジグザク列」が存在することが必要十分条件である。

定義 矢印 $f, g \in A$ が側方連結 (laterally connected) であるとは $\phi_G(f) = \phi_G(g)$ が成り立つときをいう。

側方連結性は矢印の集合 A 上 $f R_{LC} g \Leftrightarrow \phi_G(f) = \phi_G(g)$ で定義される同値関係 R_{LC} を引き起こす。以下の定理が成り立つ[11]。

定理 R_{LC} はすべての双対的視点から引き起こされる矢印の集合 A 上の同値関係の中で最小で

ある。

つまり、定理の意味で変換 L によって表される双対的視点は標準的である。この定理の証明は圏論という数学の枠組みの中でより一般の形でなされるがその内容を紹介することは本稿の目的ではないので省略する。

以上で、「モノをプロセスとして読む」双対的視点から引き起こされるネットワークの標準的構造として側方連結性を得た。次節と次々節では側方連結性の応用について述べる。

3. ネットワークのコンフィギュレーションモデルにおける側方連結性

ネットワークのコンフィギュレーションモデルとは、与えられた頂点の度数分布を持つすべての可能なネットワークの集合から一様にランダムに選ばれたネットワークのことである[17]。より簡潔には、与えられた頂点の度数分布を持つランダムネットワークのことである。本節ではコンフィギュレーション上のパーコレーションを考え、側方連結性と従来の強連結性、弱連結性との比較を行う。

連結性は通常頂点の集合上で考えられるが、本稿では矢印の集合上で考える。何故ならば、側方連結性は頂点の集合上で考えると同値関係にならず、連結成分を定義することができなくなるからである。二つの矢印 f, g が弱連結であるとは f から出発して矢印をネットワーク上で矢印の向きを無視して辿って g まで到達できるときをいう。二つの矢印 f, g が強連結であるとは f から出発して矢印をネットワーク上で矢印の向きに沿って辿って g まで到達でき、かつ g から出発して同様に f まで到達できるときをいう。側方連結性と同様に弱連結性、強連結性ともに与えられたネットワークの矢印の集合上の同値関係となり、連結成分が定義される。

コンフィギュレーションモデル上のパーコレーションは一般に母函数法を用いて解くことができる[6, 17]が、ここでは母函数法の説明は省略し、計算の結果のみを紹介するに留める。母函数法を用いることで、例えば、ジャイアントコンポーネント(ネットワーク全体の中で正の割合を占める連結成分)が出現する条件を求めることができる。また、摂動に対するネットワークの頑健性も議論することができる[21]。以下では、ジャイアントコンポーネントが出現する条件および頂点をランダムに取り除いたときにジャイアントコンポーネントが存続する確率に関する結果を紹介する。

$P(k_i, k_o)$ を、ネットワークの頂点の入力次数が k_i 、出力次数が k_o である確率とする。ある頂点を始点とする矢印は必ずネットワーク内の頂点

を終点として持つので入力次数の平均値と出力次数の平均値は等しい。つまり、 $\langle k_i \rangle = \langle k_o \rangle$ である。 $z = \langle k_i \rangle = \langle k_o \rangle$ とおくことにする。ここで、一般に k_i, k_o の函数 $f(k_i, k_o)$ に対してその平均値 $\langle f(k_i, k_o) \rangle$ は、 $\langle f(k_i, k_o) \rangle = \sum_{k_i, k_o} f(k_i, k_o) P(k_i, k_o)$ で定義した。また、 $k = k_i + k_o$ とおくことにする。このとき、ジャイアントコンポーネントの存在条件は弱連結性、強連結性、側方連結性で以下ようになる。

- (i) 弱連結性： $\langle k^2 \rangle > 2\langle k \rangle$
- (ii) 強連結性： $\langle k_i k_o \rangle > z$
- (iii) 側方連結性： $(\langle k_i^2 \rangle - z)(\langle k_o^2 \rangle - z) > z^2$

弱連結性、強連結性に対する条件は、頂点の集合上で計算された従来の結果と一致している[6, 17]。これら三つの条件の間の相互関係に関しては以下の命題が成り立つ。

命題 (1) 条件(ii)が成り立てば条件(i)が成り立つ。
(2) 条件(iii)が成り立てば条件(ii)が成り立つ。
(3) 条件(ii)と条件(iii)の間にはアプリアリな論理的依存関係は存在しない。

このうち(1)と(2)は自明であるが、(3)に関しては具体的に反例を挙げることで示すことができるが、ここではこれ以上立ち入らない。重要なことは側方連結性が強連結性とは異なるネットワークの側面を捉えることができるということが具体的に示されたということである。

次に、ジャイアントコンポーネントの摂動に対する頑健性に関する結果を述べる。特にここではネットワークから任意の頂点をランダムに確率 p で取り除いたとき、ジャイアントコンポーネントが存続する条件を求める。 $P'(k_i', k_o')$ を上述のように摂動を加えた後の有効な頂点の度数分布とすると

$$P'(k_i', k_o') = \sum_{k_i, k_o} P(k_i, k_o) \binom{k_i}{k_i'} (1-p)^{k_i - k_i'} p^{k_i - k_i'} \binom{k_o}{k_o'} (1-p)^{k_o - k_o'} p^{k_o - k_o'}$$

と書ける[21]。この等式を用いて上述の条件(i), (ii), (iii)を書き換えると

- (i') 弱連結性： $1-p > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$
- (ii') 強連結性： $1-p > \frac{z}{\langle k_i k_o \rangle}$
- (iii') 側方連結性： $1-p > \frac{z}{\sqrt{(\langle k_i^2 \rangle - z)(\langle k_o^2 \rangle - z)}}$

となる。 p_c をジャイアントコンポーネントが存続する臨界確率とする。つまり、条件(i'), (ii'), (iii')の不等式が等式となる p の値である。 $p_c = 1$ であることはネットワークが摂動に対して非常に頑健であることを示す。何故ならば確率 1 で頂点を取り除いてもジャイアントコンポーネント

が存続するということから、頂点の入出力次数の平均値が有限、つまり $z < \infty$ のときを考えよう。このとき、 $p_c = 1$ となるための条件は、(i') では $\langle k^2 \rangle$ が発散すること、(ii') では $\langle k, k_o \rangle$ が発散すること、(iii') では $\langle k_i^2 \rangle$ もしくは $\langle k_o^2 \rangle$ のいずれかが発散すること、となる。これから、 $z < \infty$ のときには、側方連結性の意味で $p_c = 1$ となることと弱連結性の意味で $p_c = 1$ となることが同値であることが分かる。この結果は側方連結性が捉えているネットワークの有向グラフとしての構造とネットワークの頑健性の間に何らかの関連があることを示唆している。しかし、コンフィギュレーションモデルでは頂点の入出力次数には頂点間で相関は無いという仮定を置いている。現実のネットワークには一般に頂点間で入出力次数に相関がある[5]ことを考慮すると、上記の結果を現実のネットワークにおいて解釈するには注意が必要である。

4. 線虫の神経ネットワークにおける側方連結性

本節では線虫 *C. elegans* の神経ネットワークにおける側方連結性の意義を調べる。特にニューロンの機能による矢印の分類と側方連結性による矢印の分類の間の類似性を定量的に評価した。以下では White ら[24]によって調べられ、Oshio ら[18]によってデータベース化されたデータを用いた。

C. elegans の神経ネットワークは 302 個のニューロンからなる。個々のニューロンを頂点とし、あるニューロンから別のニューロンに化学シナプス結合が存在するときに矢印を引いてネットワークを構成した。矢印には化学シナプス結合の個数で重みを付けた。重みに対する閾値を導入することで、各閾値の値に応じてその閾値以上の重みを持つ矢印だけを残したネットワークを考えることができる。

ニューロンには感覚ニューロン、介在ニューロン、運動ニューロンの三つの機能タイプがあり、各矢印の始点、終点のニューロンの機能タイプに応じて矢印を九つの類へと分類することができる。そこで、閾値を変化させたときにこの機能による矢印の分類と側方連結性による矢印の分類がどの程度類似しているのかを、集合上の二つの分類の類似度を測る指標を用いて調べた。用いた指標は修正ランド指数(ARI)[13, 19]および修正正規化相互情報量(AMI)[22, 23]の二つである。

結果を Fig.5 に示す。Fig.5(a) の上三角および下三角の凡例は、対照実験として矢印をランダムに張り替えた場合の計算結果であり、1000 試行の平均値である。また、エラーバーは標準偏差を示している。Fig.5(b) は ARI, AMI の Z スコア(対照実験からのずれの指標、 X という量の現実のネットワークでの値を X_{real} 、矢印をランダムに張り替

えたネットワークでの値の平均値を $\langle X_{rand} \rangle$ 、標準偏差を σ とすると X に対する Z スコアは $z_x = \frac{X_{real} - \langle X_{rand} \rangle}{\sigma}$ で定義される)を示す。閾値が 4 のときにいずれの指標の Z スコアも最大となり、値が 10 を超えている。このことは、閾値が 4 のときの *C. elegans* の神経ネットワークの機能的構造と側方連結性との間に何らかの関連があることを示唆している。ただし、Fig.5(a) の縦軸の最大値は 1 であり、類似度の絶対値としてはそれほど高くはないことに注意する必要がある。

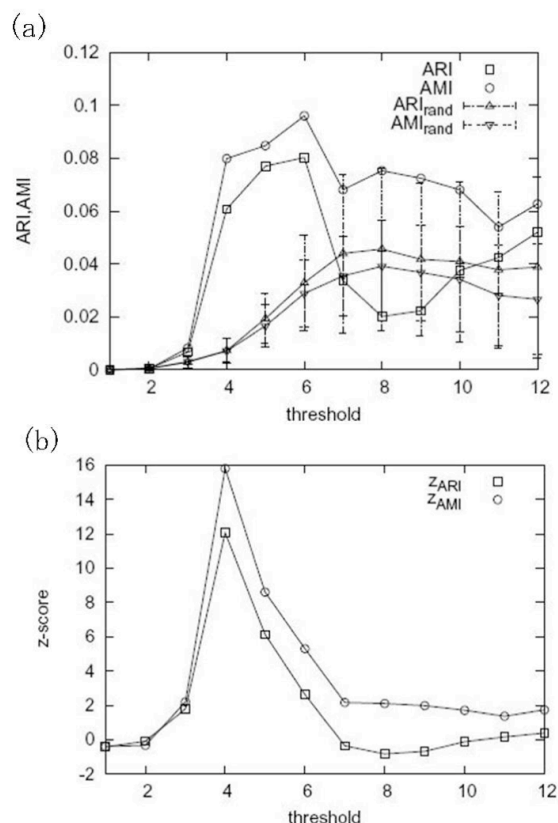


Fig.5 (a) Comparison between clustering of arcs by lateral connectedness and clustering of arcs based on functions of neurons in a neuronal network of *C. elegans*. For each threshold value we constructed a network in which an arc between two nodes exists if and only if the number of chemical synapses from one node to the other node is more than or equal to the threshold value. We calculated similarity indices ARI(adjusted Rand index) and AMI(adjusted normalized mutual information) between the two clusterings for each threshold value. ARI_{rand} and AMI_{rand} are averages for 1000 randomized networks. Error bars indicate standard deviations. (b) Z-scores for ARI and AMI. Both of them take their maximum values at threshold 4 and their values are extremely high (more than 10).

5. おわりに

本稿では「モノをプロセスとして読む」双対的視点を複雑ネットワークへと導入し、その数学的定式化と応用に関して紹介した。まず、双対的視

点から導かれるネットワークの標準的な構造としての側方連結性を導入した。次に、コンフィギュレーションモデルにおいて側方連結性と従来の強連結性・弱連結性を比較し、特に側方連結性とネットワークの頑健性との関連を示唆する結果を得た。また、線虫 *C. elegans* の神経ネットワークの機能的構造と側方連結性との関連を示唆する結果を紹介した。本稿で紹介した内容[11]は筆者により始められたばかりの研究であり、その数学的側面および応用の両者に関していまだに発展途上の段階にある。ネットワークに対して現実的な見方をするだけでなく、一旦抽象的な双対的視点へと飛び上がり、そこからまた下降してくることで現実世界の複雑ネットワークのさらなる深い理解が得られるにちがいない。

謝辞

本研究は、JST 戦略的創造研究推進事業さきがけの一環として行われたものである。

References

1. Ahn, Y.-Y., Bagrow, J. P. and Lehmann, S. Link communities reveal multiscale complexity in networks, *Nature* 466, 761-765 (2010).
2. Alon, U. *Introduction to Systems Biology: Design Principles of Biological Circuits*, CRC Press, Boca Raton, 2006.
3. Albert, R. and Barabasi, A.-L. Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74, 47-97 (2002).
4. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.-U. Complex networks: Structure and dynamics, *Phys. Rep.* 424, 175-308 (2006).
5. Boguna, M. and Serrano, A. Generalized percolation in random directed networks, *Phys. Rev. E* 72, 016106 (2005).
6. Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. Giant strongly connected component of directed networks, *Phys. Rev. E* 64, 025101(R) (2001).
7. Evans, T. S. and Lambiotte, R. Line graphs, link partitions, and overlapping communities, *Phys. Rev. E* 80, 016105 (2009).
8. Haruna, T. *Algebraic Theory of Biological Organization*: pp. 103, Doctoral Dissertation, Kobe University, 2008.
9. Haruna, T. Being Arranged in Advance: Quantum Entanglement and Biological Feedback, pp.220-226, in Bullock, S., Noble, J., Watson, R. and Bedau, M. A. Eds., *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, MIT Press, Cambridge, MA, 2008.
10. Haruna, T. An Application of Category Theory to the Study of Complex Networks, *Int. J. Comp. Anti. Sys.* 23, 146-157.
11. Haruna, T. Global Structure of Directed Networks Emerging from a Category Theoretical Formulation of the Idea "Objects as Processes, Interactions as Interfaces", pp. 310-317, in T. Lenaerts et al. Eds., *Advances in Artificial Life, ECAL 2011, Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*, 2011.
12. Haruna, T. and Gunji, Y.-P. Duality between decomposition and gluing: A theoretical biology via adjoint functors, *BioSystems* 90, 716-727 (2007).
13. Hubert, L. and Arabie, P. Comparing Partitions, *J. Classification* 2, 193-218 (1985).
14. MacLane, S. and Moerdijk, I. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer-Verlag, New York, 1992.
15. MacLane, S. *Categories for the Working Mathematician*, 2nd edition, Springer-Verlag, New York, 1998.
16. Newman, M. E. J. The structure and function of complex networks, *SIAM Review* 45, 167-256 (2003).
17. Newman, M. E. J., Strogatz, S. H. and Watts, D. J. Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* 64, 026118 (2001).
18. Oshio, K., Iwasaki, Y., Morita, S., Osana, Y., Gomi, S., Akiyama, E., Omata, K., Oka, K. and Kawamura, K. Database of Synaptic Connectivity of *C. elegans* for Computation, Technical Report of CCeP, Keio Future, No.3, Keio University, (2003).
19. Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods, *J. Am. Stat. Assoc.* 66, 846-850 (1971).
20. Rosen, R. The representation of biological systems from the standpoint of the theory of categories, *Bull. Math. Biophys.* 20, 317-341 (1958).
21. Schwartz, N., Cohen, R., ben-Avraham, D., Barabasi, A.-L. and Havlin, S. Percolation in directed scale-free networks, *Phys. Rev. E* 66 015104(R) (2002).
22. Strehl, A. and Ghosh, J. Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions, *J. Mach. Learn. Res.* 3, 583-617 (2002).
23. Vinh, N. X., Epps, J. and Bailey, J. Information Theoretic Measures for Clusterings Comparison: Is a Correction for Chance for Necessary? in *Proc. 26th Int. Conf. Mach. Learn.*, Montreal, Canada, 2009.
24. White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*, *Phil. Trans. R. Soc. London B* 314, 1-340 (1986).