

特集 「人と環境に見る高次元データフローの生成と解析」

テンポラルネットワーク

Temporal Networks: A Short Introductory Survey

増田 直紀
Naoki Masuda

東京大学大学院情報理工学系研究科数理情報学専攻
Graduate School of Information Science and Technology, The University of Tokyo.
masuda@mist.i.u-tokyo.ac.jp, <http://www.stat.t.u-tokyo.ac.jp/~masuda/>

Keywords: network, burst, sociophysics, point process.

1. テンポラルネットワークとは

ネットワークは頂点と枝の集合からなり、グラフ理論の指すグラフと同値である。人間関係、インターネット、Web、遺伝子の制御関係など、さまざまな対象がネットワークとして解析され、その理解がここ 15 年程度で急速に進んだ。それらの発展については、いくつもの成書がある ([増田 10] や、そこにある解説付きの参考文献を参照。より最近の書籍の例には日本語では [増田 12, 安田 10, 安田 11], 英語では [Cohen 10, Newman 10] などがある。また、本誌に掲載された総説に [増田 08] がある)。

現実のネットワークのデータは、しばしば頂点や枝の数において巨大であり、その意味で、本特集が対象とするマッシュデータであるといえる。しかし、本稿では、頂点や枝の数以外の意味でマッシュデータの種類とみなすことができる、近年取得されている種類のデータに焦点を当てる。それは、時間情報を含むネットワークである。従来、ネットワーク科学は時間情報をあまり詳細に扱ってこなかったと思われるが、データの整備とともに、従来のネットワークの見方が拡張を迫られているのである。

図 1 を考える。左図は、A, B, C, D の四人についての、相互作用イベントの時系列を表す。相互作用イベントとは、メールの送受信や会話など、個人間のやり取りを表す。この図では、A と B は 8:50 に 1 回会話を行っている。イベントの持続時間は、本稿では簡単のために考えない。左図をネットワーク風に表すと、右図のようになる。従来の静的なネットワークと異なるのは、時刻情報が付加されていることである。リンクは一瞬だけ使われるのである。

このような、リンクに時刻情報が付加されたネットワークのデータは、テンポラルネットワーク (略して、テンポラルネット) と総称される。近年のデータの蓄積に伴って、テンポラルネットの解析が急速に進められるようになってきている。総説も出された [Holme 11]。拙著でも簡単に紹介した [増田 12, 第 8 章]。時間軸があるために、頂点や枝が非常に多くはない場合でも、時間

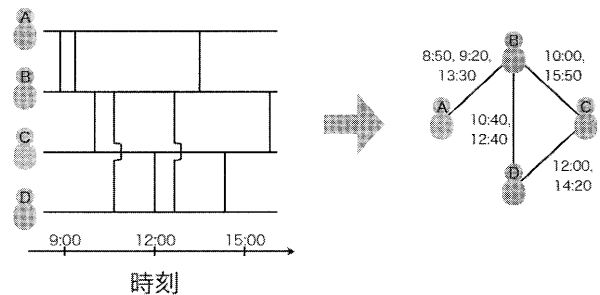


図 1 リンクは一瞬だけ使われる

方向に大量の相互作用イベントがあれば、テンポラルネットはマッシュデータの一例と呼べるだろう。そして、このようなデータの解析手法は、まだ確立していない。ネットワーク科学は、統計力学的な手法 (平均場近似や大規模な数値計算など) を用いる人々が最も主体になってここまで行われてきたように思われる (もちろん、計算機科学や機械学習の研究も今では多くある)。しかし、テンポラルネットの理解には、統計力学的な手法に加えて、確率過程 (特に点過程)、統計、機械学習等の見解も欠かせないと思われる。

2. テンポラルネットワークを考える必要性

時刻情報が入手できるようになったからといって、それを用いることが有効な解析手段を与えるとは限らない。時刻情報をあえて捨てて従来型の静的なネットワークとして単純化して物事を解析するのも、一つの方針である。テンポラルネットを考えることは必要なのだろうか? その必要性を示すために、テンポラルネットに特徴的な二つの事実を説明する。

第一に、イベント時刻の系列は間欠的であることが多い [Barabási 05, Barabási 10, Castellano 09, Vázquez 06]。間欠的とは、イベント間隔の長いときと短いときの差が極端であることであり、図 2 がその例である。図 2 は、一人についてのイベント時系列を示したものである。ある相手を固定した上でのイベント時系列の場合も、相手を問わずに混ぜたうえでのイベント時系列の場合も、イベント間隔が間欠的であることは普遍的に起こる。

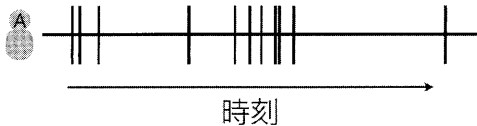


図2 イベント間隔は、間欠的であることが多い

より定量的に述べると、イベント間隔の分布がべき則に代表されるような長い裾をもつ分布であることが非常に多い。そして、その生成モデルもいくつか提出されている [Barabási 05, Castellano 09]。感染症、意見伝搬、ゲームなどの社会ダイナミクスの従来のモデルでは、この間隔が指数分布に従うことを暗黙に仮定してきた。すなわち、確率過程で言えば、イベント時系列はポアソン過程に従って発生することが仮定されてきた。シミュレーションにおいて、意見を変更するといった何らかの状態変化を起こすエージェント（頂点）や相互作用を行う二人のエージェント（枝）を、毎回ランダムに選択することは標準的だ。この場合も、イベント間隔が頂点ないし枝ごとに独立な指数分布に従うと仮定されていると思ってよい。指数分布の仮定を置くのは、シミュレーションを行うときに独立な状態更新に対応するので自然であるということと、数学的に扱いやすいことに起因する。しかし、実データは、指数分布の仮定を支持しないことが多いのである。

第二の事実を説明するために、図3を考える。左図は、テンポラルネットワークの一表現である。時刻情報を無視して静的なネットワークとして考えると、右図となる。右図では、AとCは対称な位置にある。しかし、左図によると、Aの発した情報はBを介してCに届くものの、Cの発した情報はAには届かない。これは、枝ABと枝BCの太さ（=イベントの回数）が異なることが理由ではない。相互作用の時間的な方向性が理由である。個々のイベントには方向性がないが、イベントの時空間系列には、イベントの全体的な流れが存在し得るのである。図3は極端な例だが、この非対称性は、テンポラルネットワーク一般について当てはまることである。

このように、テンポラルネットワークの実データの特徴は、従来のモデルの仮定と異なることが多い。このような差異が明らかにされてくるにつれて、ネットワーク上で起こるダイナミクスの理解も変更が迫られている。例え

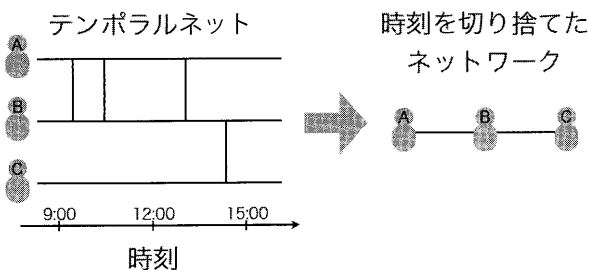


図3 テンポラルネットワークには、イベントが無向でも、イベントの全体的な流れが存在し得る

ば、第一の事実に対応するために、イベント間隔を指数分布でなく裾の長い分布に変更して解析を行う。あるいは、第一と第二の事実に対応することを意図して、実データそのもののイベント間隔の時系列を用いて、ダイナミクスの様子を調べる。例えば、SIモデルやSIRモデルという感染症の伝播を表す代表的なモデルにおいて、テンポラルネットワークの実情を考慮すると、感染の広がり方が静的なネットワークの場合よりも遅くなるなどの影響が出ることが示唆されている [Iribarren 09, Isella 11a, Isella 11b, Karrer 10, Karsai 11, Min 11, Miritello 11, Rocha 11, Vazquez 07]。また、集団合意形成のモデルにおいても、イベント間隔の分布は全体の合意が達成されるまでの時間などに影響する、という数値計算結果がある [Fernández-Gracia 11, Stark 08, Takaguchi 11]。

テンポラルネットワークの解析手法を開発することは、現在のネットワーク研究のチャレンジの一つであると思われる。このようなデータを眺める視点は現時点では無数にあり、定説はないと思う。データ解析の有効な切り口を一つずつつくり、また、提案された一つの手法を自分の手持ちのデータだけでなくさまざまなテンポラルネットワークのデータに適用して有効性を確かめる、といった作業が現時点では有意義であろう。以下の章では、解析手法のうち二つを紹介する。代表的な二つという意味ではなく、それなりに多数ある解析手法のうちの二例という位置付けである。

3. 会話相手の予測可能性

企業のオフィスで働く社員のコミュニケーション行動を考える。昼休みには同僚とともに昼食に出かけて会話を楽しむかもしれない。また、職務によっては、ある部署の人から渡された案件をいつも同じ次の担当者へ渡すために会話をすることもあるかもしれない。各個人に注目したもでの会話相手の順序は、企業組織などにおける情報の流れを決める一つの要素だと考えられる。このことは、組織に限らず当てはまるだろう。本章で紹介する研究では、実際のデータに対して、人の会話相手の選択について、規則性の度合いの定量化を行った。

図2を、Aの関わるイベント（本章では会話と呼ぶ）の相手は固定されていない状況を表すものとみなす。このとき、会話相手の順番はランダムであろうか？ AにはBとCという二人の会話相手がいると仮定する。すなわち、図2の各会話はBまたはCと行われたとする。そして、AはBとよりも2倍多くCと会話している状況を考える。すると、任意の時刻において、それまでの会話相手の履歴によらずに次の会話相手がBである確率が1/3、Cである確率が2/3であるという仮定が、最も単純なランダム性の仮定であろう。

著者は、日立製作所中央研究所と自身の学生との共同研究によって、会話相手の予測可能性を調べた

[Takaguchi 11]. 分析には日立製作所ワールドシグナルセンタが収集したデータを利用した。データ収集に利用されたビジネス顕微鏡の Web サイト [Biz] も参照されたい。被験者は、とある日本企業のオフィスで働く社員の方々だ。就業中に、胸の前に名札型のセンサを付けてもらう。そして、センサ同士の空間的な距離などに基づいて個人間の会話イベントが検出される。誰が誰といつ会話をしたかという情報が約 160 人、約 2 か月半にわたって得られ、それを活用した。また、イベントの詳細な時刻は無視し、会話相手の順番（これも、テンポラルネットの重要な要素である）のみに着目して解析を行った。

今回誰と会話したかを知ると、次の会話相手を予測することがそれなりの程度できることがわかった。エントロピーを用いて、個々の人について、予測可能性を定量化した。現在の会話相手で条件を付けない次の会話相手の分布のエントロピーと、現在の会話相手で条件付けた際の次の会話相手の分布のエントロピーの差が大きい場合を考える。このときは、現在の会話相手を知ることによって次の会話相手についての不確実性が減少しているので、この人は会話相手の予測可能性が高い人であるとみなす。エントロピーの差が小さい場合は、予測可能性が低い人であるとみなす。

この指標を用いて、二つのデータセットについて、十分な会話イベント数がある個人について調べたところ、ほぼすべての人について、有意な予測可能性があることがわかった。実際には、予測可能性の大部分は、今回話した相手と次回も話しやすいというデータの特性から来ている。ただし、その要素を取り除いたうえで解析を行っても、ほとんどの被験者について予測可能性がある程度残存することがわかった。

次に、予測可能性には個人差がある。この個人差と、個人のネットワーク上での位置を関係付けることを試みた。ここでのネットワークは、静的なネットワークのことである。枝の太さは、計測期間中に二人の間に起こった会話イベント数を表す。予測可能性が高い人は、周りに三角形をあまりもたず、また、細い枝をもつ傾向があることが明らかにされた。このような人は、直観的には、ネットワークの異なるコミュニティ（内部では枝が密で外部へ向かっては枝が疎であるようなネットワークの部分のこと [Fortunato 10]) を結ぶ位置にいることが多い。実際に、細い枝は、一つのコミュニティ内の二人を結ぶよりも複数のコミュニティを結ぶ傾向があることを確認した。一方、予測可能性が低い人は、周りに三角形や太い枝を持ちやすく、そのような人は、コミュニティの外部に届く枝をあまりもっていないことが多い。

4. 動的な中心性

中心性とは、ネットワークの中でどの頂点が重要であるかを測る指標である。何をもって重要であるとするか

は、ネットワーク上で何が起こることを想定するかに依存する [増田 11, 増田 12]。したがって、中心性の指標は多数存在する。とはいえ、静的なネットワークに対して実際によく用いられている中心性の指標は、5 種類程度であると大ざっぱにはいえる [増田 12]。例えば、次数（枝で直接つながっている相手の数）は、最も単純かつ基本的な中心性指標である。ハブ（次数が非常に大きい頂点）がネットワークの中心であるという基準である。ほかの中心性では、媒介中心性、ページランクなどが有名である [増田 10]。

テンポラルネットにおける中心性は、静的なネットワークにおける中心性とは一般的に異なる。テンポラルネットに対する中心性の考え方には二つがあると思われる。一つ目は、イベントの時刻情報を用いつつも、静的なネットワークの場合と同様に、個々の頂点には時間的に変化しない一つの値を割り当てるという考え方である [Grindrod 11, Pan 11, Tang 10]。例えば、静的なネットワークにおける媒介中心性とは、頂点同士の通信がネットワーク上の最短経路に沿って起こると仮定した場合に、各頂点が、自分以外の頂点对を結ぶ最短経路の上にいる程度を表す。しかし、テンポラルネットにおいては、時刻情報を捨て去って得られた静的なネットワークにおける最短経路が実現不能かもしれない。実現不能とは、時間関係の辻褄があっていないことを指す。例えば、4 人のネットワークで、A と B が午前 9 時に会話し、B と C が午後 3 時に会話し、C と D が午前 10 時に会話したとする。すると、静的なネットワークにおいては、A と D を結ぶ最短経路は、A から B, C とたどって D に到達する道である。枝の方向性は考えないとすると、D から A への最短経路も同じである。B と C は、A と D を結ぶ最短経路上にいる、ということになる。しかし、テンポラルネットでは、ほかに会話イベントがないとすると、A と D は結ばれない。

二つ目は、個々の頂点の中心性の値が、時間的に一定ではなく、テンポラルネットの観測とともに動的に動くという考え方である。スポーツのランキングを例にとって説明する。スポーツの選手やチームのランキングは、しばしば対戦相手を方向付きのリンクで結んだネットワークにおける中心性の問題として捉えることができる。日本のプロ野球や J リーグならば、すべてのチームが総当たりで同じ対戦回数だけ対戦するので、ネットワークとして考える必要はない。勝率や勝ち点によって、大きな問題なく強さを判定できる。しかし、世界の国・地域別のサッカーや、テニスなどの個人スポーツでは、状況が異なる。すなわち、長い期間にわたって考えたとしても、各個人が対戦する相手の集合は異なる（以下の議論は、チームスポーツの場合にも同様に適用できる）。この場合は、勝率や勝ち点で個人の強弱を判断できない。なぜなら、弱い相手と対戦しやすい個人は勝ちを稼ぎやすく、逆も然りだからである。このような状況において

統計的な手法で個々の強さを推定する手法は、昔からある。そして、アメリカの大学におけるアメリカンフットボール [Park 05] や男子のプロテニス [Radicchi 11] などに、ネットワークに基づくランキング手法が応用されている。

これらの先行研究では、選手の静的なネットワークに対して中心性を定めている。実際には、選手の強さは時々刻々と変化する。最初は弱く、ある段階で強さのピークを迎え、その後は、引退に向かって弱くなっていくことが典型的だろう。このような中心性の時間依存性を捉えるにはどうすればよいだろうか。

1回のゲームの結果は、テンポラルネットワークの一つのイベントとみなすことができる。すると、問題は、テンポラルネットワークに対して、時間依存の中心性を定義することとなる。著者は、学生とともに、静的な対戦ネットワークについての先行研究 [Park 05] をテンポラルネットワークの場合に拡張することによって、テニスの選手の動的なランキングを行った [茂木 12, Motegi 12]。具体的には、各選手のスコア（中心性の値）は、試合のない期間は、時間に対して指数的に減衰すると仮定した。そして、新しく誰かに勝つ度に、スコアがある値だけ加算される（負けの場合も同様）。指数的減衰の仮定によって、試合の結果と時刻が逐一追加されるごとに各選手のスコアを簡単な数式に基づいて更新することが可能になる。動的なランキング手法は、従来手法よりも試合結果の予測精度（ある時刻までの試合結果を知ってスコアを算出したうえで、次の試合で、高いスコアのほうの選手が低いスコアのほうの選手に勝つ割合）が若干高いことがわかった。指数的減衰の時定数などのパラメータ値に関するある程度の頑健性（値を変えてもランキング結果がさほど変化しないこと）も確認した。また、期待されたとおり、選手の強さが時間とともに大きくなり、ピークを迎え、やがて小さくなる（徐々に弱くなる場合も、まだ強いときに引退する場合も含まれる）ことを示した。

スポーツを例にとって説明したが、動的な中心性が重要になる状況は数多い。脳では、タスクや情報処理の段階に応じて、同じニューロンが中心的であったりなかったりするかもしれない。金融機関のネットワークでは、資金の流れが時々刻々と変化することによって、中心である金融機関が変化していくかもしれない。

5. ま と め

頂点間の相互作用イベント時系列というデータに対して、テンポラルネットワークという視点を紹介した。イベントの相手の予測可能性（3章）と中心性（4章）という二つの話題に絞って説明を行った。これらは、テンポラルネットワークというマッシュデータを要約する手法の二つ、と位置づけられる。例えば、中心性とは、複雑なネットワーク構造や時間情報に起因する多大であろう情報量を、

頂点の得点というスカラ値や順位付けに要約することだ。そのように詳細を捨象すると、理解が良くなったり、何らかの操作手法の設計に有用だったりすることがしばしばあると思われる。実際には、ほかにもさまざまなテンポラルネットワークの解析手法がある [Holme 11]。今後も、いろいろな研究領域（特に、統計や機械学習）の研究者の参画を通じて、より強力な解析手法が発展することが期待される。

謝 辞

本原稿は科研費（23681033 および 20115009）の助成を受けたものである。本原稿を読んでコメントを下さった高口太郎氏（東京大学）に御礼申し上げる。

◇ 参 考 文 献 ◇

- [Biz,] <http://www.hitachi-hitec.com/jyouhou/business-microscope/>
- [Barabási 05] Barabási, A. L.: The origin of bursts and heavy tails in human dynamics, *Nature*, Vol. 435, pp. 207-211 (2005)
- [Barabási 10] Barabási, A. L.: *Bursts—The Hidden Pattern Behind Everything We Do*, Dutton, New York (2010)
- [Castellano 09] Castellano, C., Fortunato, S. and Loreto, V.: Statistical physics of social dynamics, *Rev. Mod. Phys.*, Vol. 81, pp. 591-646 (2009)
- [Cohen 10] Cohen, R. and Havlin, S.: *Complex Networks—Structure, Robustness and Function*, Cambridge University Press, Cambridge (2010)
- [Fernández-Gracia 11] Fernández-Gracia, J., Eguíluz, V. M. and San Miguel, M.: Update rules and interevent time distributions: Slow ordering versus no ordering in the voter model, *Phys. Rev. E*, Vol. 84, 015103 (R) (2011)
- [Fortunato 10] Fortunato, S.: Community detection in graphs, *Phys. Rep.*, Vol. 486, pp. 75-174 (2010)
- [Grindrod 11] Grindrod, P., Parsons, M. C., Higham, D. J. and Estrada, E.: Communicability across evolving networks, *Phys. Rev. E*, Vol. 83, 046120 (2011)
- [Holme 11] Holme, P. and Saramäki, J.: Temporal networks, arXiv:1108.1780v1 (2011)
- [Iribarren 09] Iribarren, J. L. and Moro, E.: Impact of human activity patterns on the dynamics of information diffusion, *Phys. Rev. Lett.*, Vol. 103, 038702 (2009)
- [Isella 11a] Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., Gesualdo, F., Pandolfi, E., Ravà, L., Rizzo, C. and Tozzi, A. E.: Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors, *PLoS ONE*, Vol. 6, e17144 (2011a)
- [Isella 11b] Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., and Van den Broeck, W.: What's in a crowd? Analysis of face-to-face behavioral networks, *J. Theoretical Biology*, Vol. 271, pp. 166-180 (2011b)
- [Karrer 10] Karrer, B. and Newman, M. E. J.: Message passing approach for general epidemic models, *Phys. Rev. E*, Vol. 82, 016101 (2010)
- [Karsai 11] Karsai, a, M., Kivela, M., Pan, R., Kaski, K., Kertész, J., Barabási, A.-L. and Saramäki, J.: Small but slow world: How network topology and burstiness slow down spreading, *Phys. Rev. E*, Vol. 83, 025102 (R) (2011)
- [増田 08] 増田直紀: ネットワーク上の進化ゲーム, 人工知能学会誌, Vol. 23, No. 5, pp. 652-658 (2008)
- [増田 11] 増田直紀: ネットワークの重要人物は誰か, DIAMOND ハーバード・ビジネス・レビュー, 8月号, p. 1 (2011)
- [増田 12] 増田直紀: なぜ3人いると噂が広まるのか, 日本経済新聞出版社 (2012)

- [増田 10] 増田直紀, 今野紀雄: 複雑ネットワーク, 基礎から応用まで, 近代科学社 (2010)
- [Min 11] Min, B., Goh, K.-I. and Vazquez, A.: Spreading dynamics following bursty human activity patterns, *Phys. Rev. E*, Vol. 83, 036102 (2011)
- [Miritello 11] Miritello, G., Moro, E. and Lara, R.: Dynamical strength of social ties in information spreading, *Phys. Rev. E*, Vol. 83, 045102 (R) (2011)
- [茂木 12] 茂木 隼, 増田直紀: プロテニスプレーヤーの動的なランキング, 第8回ネットワーク生態学シンポジウム予稿集 (2012)
- [Motegi 12] Motegi, S. and Masuda, N.: A network-based dynamical ranking system, arXiv:1203.2228v1 (2012)
- [Newman 10] Newman, M. E. J.: *Networks—An Introduction*, Oxford University Press, Oxford (2010)
- [Pan 11] Pan, R. K. and Saramäki, J.: Path lengths, correlations, and centrality in temporal networks, *Phys. Rev. E*, Vol. 84, 016105 (2011)
- [Park 05] Park, J. and Newman, M. E. J.: A network-based ranking system for US college football, *J. Stat. Mech.*, P10014 (2005)
- [Radicchi 11] Radicchi, F.: Who is the best player ever? A complex network analysis of the history of professional tennis, *PLoS ONE*, Vol. 6, e17249 (2011)
- [Rocha 11] Rocha, L. E. C., Liljeros, F. and Holme, P.: Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts, *PLoS Comput. Biol.*, Vol. 7, e1001109 (2011)
- [Stark 08] Stark, H. U., Tessone, C. J. and Schweitzer, F.: Decelerating microdynamics can accelerate macrodynamics in the voter model, *Phys. Rev. Lett.*, Vol. 101, 018701 (2008)
- [Takaguchi 11] Takaguchi, T. and Masuda, N.: Voter model with non-poissonian interevent intervals, *Phys. Rev. E*, Vol. 84, 036115 (2011)
- [Takaguchi 11] Takaguchi, T., Nakamura, M., Sato, N., Yano, K. and Masuda, N.: Predictability of conversation partners, *Phys. Rev. X*, Vol. 1, 011008 (2011)
- [Tang 10] Tang, J., Musolesi, M., Mascolo, C., Latora, V. and Nicosia, V.: Analysing information flows and keymediators through temporal centrality metrics, *Proc. 3rd Workshop on Social Network Systems* (2010)
- [Vázquez 06] Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. I., Kondor, I. and Barabási, A. L.: Modeling bursts and heavy tails in human dynamics, *Phys. Rev. E*, Vol. 73, 036127 (2006)
- [Vázquez 07] Vázquez, A., Rácz, B., Lukács, A., and Barabási, A. L.: Impact of non-Poissonian activity patterns on spreading processes, *Phys. Rev. Lett.*, Vol. 98, 158702 (2007)
- [安田 10] 安田 雪: つながりを突き止める, 入門! ネットワーク・サイエンス, 光文社 (2010)
- [安田 11] 安田 雪: パーソナルネットワーク, 人のつながりがもたらすもの, 新曜社 (2011)

2012年5月8日 受理

著者紹介



増田 直紀

1998年に東京大学工学部計数工学科を卒業した後、同大学院に進学し、2002年に博士(工学)。東京大学准教授(大学院情報理工学系研究科数理情報学専攻)。ネットワーク科学、さまざまな社会行動の数理、脳の理論を研究。著書に『「複雑ネットワーク」とは何か』(共著, 講談社ブルーバックス), 『私たちはどうつながっているのか』(単著, 中公新書), 『複雑ネットワーク』(共著, 近代科学社), 『なぜ3人いると噂が広まるのか』(単著, 日本経済新聞出版社)など。