

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6251370号
(P6251370)

(45) 発行日 平成29年12月20日 (2017.12.20)

(24) 登録日 平成29年12月1日 (2017.12.1)

(51) Int. Cl.	F I		
G 0 6 F 19/12 (2011.01)	G 0 6 F	19/12	
C 1 2 M 1/00 (2006.01)	C 1 2 M	1/00	A
C 1 2 Q 1/68 (2006.01)	C 1 2 Q	1/68	A

請求項の数 20 外国語出願 (全 47 頁)

(21) 出願番号	特願2016-240912 (P2016-240912)	(73) 特許権者	500586875
(22) 出願日	平成28年12月13日 (2016.12.13)		フィリップ モリス プロダクツ エス
(62) 分割の表示	特願2014-526520 (P2014-526520)		アー
原出願日	平成24年8月24日 (2012.8.24)		スイス国 2000 ヌーシャテル ケ
(65) 公開番号	特開2017-84383 (P2017-84383A)	(74) 代理人	100078282
(43) 公開日	平成29年5月18日 (2017.5.18)		弁理士 山本 秀策
審査請求日	平成28年12月13日 (2016.12.13)	(74) 代理人	100113413
(31) 優先権主張番号	61/527, 946		弁理士 森下 夏樹
(32) 優先日	平成23年8月26日 (2011.8.26)	(74) 代理人	100181674
(33) 優先権主張国	米国 (US)		弁理士 飯田 貴敏
		(74) 代理人	100181641
			弁理士 石川 大輔
		(74) 代理人	230113332
			弁護士 山本 健策

最終頁に続く

(54) 【発明の名称】 トポロジーネットワーク攪乱を特徴づけるためのシステムおよび方法

(57) 【特許請求の範囲】

【請求項1】

生物系のネットワークモデルを特徴付けるためのコンピュータ化された方法であって、ネットワークモデルを識別することであって、前記ネットワークモデルは、前記生物系を表し、前記ネットワークモデルは、複数の生物学的実体を表す複数のノードと、複数のノード間の関係を表す複数のエッジとを含む、ことと、

プロセッサで、作用物質に対する前記生物系の応答に対応する一組の処置データと一組の制御データとを受け取ることと、

前記複数のノードのサブセットに対する前記処置データと前記制御データとの間の差を決定することと、

前記差から、各それぞれのエッジに対して、遷移確率を決定することであって、前記遷移確率は、前記それぞれのエッジによって接続された、対応する第1のノードから対応する第2のノードへの遷移の尤度を表す、ことと、

前記対応するノードへのランダムウォークの予想される訪問の回数を決定することによって、対応するノードに対する遷移確率から、前記対応するノードに対する中心度値を取得することと

を含む、方法。

【請求項2】

前記差は、前記複数のノードの活性が前記生物系に対する前記作用物質の攪乱によって影響を受ける程度を表し、前記中心度値は、前記ネットワークにおける前記対応するノ

ドの相対的な重要度を表す、請求項 1 に記載の方法。

【請求項 3】

前記複数のノードに対する攪乱指標を決定することをさらに含み、前記攪乱指標は、前記処置データと前記制御データとの間の前記差を表し、前記複数のエッジに対する遷移確率は、前記攪乱指標から決定される、請求項 1 ~ 2 のいずれかに記載の方法。

【請求項 4】

前記遷移確率は、前記攪乱指標の一次関数である、請求項 3 に記載の方法。

【請求項 5】

特定のノードに対する少なくとも 1 つの攪乱指標は、前記特定のノードから下流のノードの測定された活性の一次結合として計算される、請求項 3 ~ 4 のいずれかに記載の方法

10

【請求項 6】

定常状態において前記生物学的実体を表すノードへのランダムウォークの訪問の確率を表す平衡確率を、前記生物学的実体を表すノードに対して決定することをさらに含む、請求項 1 ~ 5 のいずれかに記載の方法。

【請求項 7】

前記中心度値は、前記平衡確率に基づいて計算される、請求項 6 に記載の方法。

【請求項 8】

前記差は、前記複数のノードのサブセットにおける対応するノードでの前記処置データと前記制御データとの間の倍率変化値である、請求項 1 ~ 7 のいずれかに記載の方法。

20

【請求項 9】

第 2 のノードに対する前記処置データと前記制御データとの間の差に関する第 1 のノードに対する中心度値の偏導関数を計算することをさらに含み、前記偏導関数は、前記ネットワークモデルに対するトポロジー感度尺度を表す、請求項 1 ~ 8 のいずれかに記載の方法。

【請求項 10】

前記偏導関数を計算することは、前記第 1 のノードの中心度値の変化に及ぼす第 2 のノードに対する前記処置データと前記制御データとの間の差の変化の影響を決定することを含み、請求項 9 に記載の方法。

【請求項 11】

前記遷移確率は、前記対応する第 1 のノードから出て行くエッジの数に基づいて決定される、請求項 1 ~ 10 のいずれかに記載の方法。

30

【請求項 12】

前記遷移確率は、前記対応する第 2 のノードが、前記複数のノードのサブセット内のいずれかのノードに直接的に接続されているかどうかに基づいて決定される、請求項 1 ~ 11 のいずれかに記載の方法。

【請求項 13】

前記対応する第 2 のノードは、前記複数のノードのサブセット内のいずれのノードにも直接的に接続されておらず、前記遷移確率は、前記対応する第 1 のノードから出て行くエッジの数によって除算された数に比例する、請求項 12 に記載の方法。

40

【請求項 14】

前記対応する第 2 のノードが、前記複数のノードのサブセット内の少なくとも 1 つのノードに直接的に接続されている場合、前記遷移確率は、付加的な因子と増幅された攪乱指標との間の合計に比例し、前記対応する第 1 のノードから出て行くエッジの数によって正規化され、前記増幅された攪乱指標は、前記対応する第 2 のノードにおける前記処置データと前記制御データとの間の増幅された差を表す、請求項 12 ~ 13 のいずれかに記載の方法。

【請求項 15】

スペクトル変換ベクトル上への中心度値の射影を、前記中心度値と、前記ネットワークモデルに対する攪乱の影響を表す前記ネットワークモデルとに基づいて、決定することを

50

さらに含む、請求項 1 ~ 14 のいずれかに記載の方法。

【請求項 16】

前記中心度値の射影を決定することは、前記中心度値をフィルタリングすることを含む、請求項 15 に記載の方法。

【請求項 17】

前記生物系の攪乱を表す前記ネットワークモデルに対するスコアを生成するために前記中心度値を収集することをさらに含む、請求項 1 ~ 16 のいずれかに記載の方法。

【請求項 18】

前記スコアは、スカラー値である、請求項 17 に記載の方法。

【請求項 19】

前記中心度値を収集することは、前記中心度値の一次結合を計算することを含む、請求項 17 ~ 18 のいずれかに記載の方法。

【請求項 20】

前記中心度値を収集することは、前記中心度値のスペクトル変換の一次結合を計算することを含む、請求項 17 ~ 19 のいずれかに記載の方法。

【発明の詳細な説明】

【背景技術】

【0001】

背景

人体は、長期間にわたって重大な健康危険要因となりうる潜在的に有害な作用物質への曝露によって常時攪乱されている。これらの作用物質への曝露で、人体内部の生物学的機構の正常な機能が損なわれる可能性がある。これらの攪乱 (p e r t u r b a t i o n) が人体に及ぼす作用を理解し、定量化するために、研究者らは、生物系が作用物質への曝露に応答する機構を研究している。いくつかのグループが i n v i v o 動物試験法を広範に利用してきたが、動物試験から得られる応答がヒト生物学に外挿されうるかどうかに関して疑念がある。他の方法として、ヒトの志願者での臨床研究を通じて危険性を評価することが挙げられるが、i n v i t r o の細胞および組織ベースの方法は、これに対応する動物ベースの方法に対する完全な、または部分的な代替方法として一般的な容認を受けているが、これらの方法は限られた価値を持つ。i n v i t r o 法は、細胞および組織の機構の特定の態様に焦点をあわせるものであるため、生物系全体に生じる複雑な相互作用を常に考慮するわけではない。

【0002】

この10年間のうちに、従来用量依存的な効力および毒性アッセイと併せた核酸、タンパク質、および代謝物レベルのハイスループット測定が、多くの生物学的過程の作用機構を解明するための手段として登場した。研究者らは、これらの異なる測定結果からの情報を科学文献からの生物学的経路に関する知識と組み合わせて意味のある生物学的モデルを構築することを試みた。この目的のために、研究者らは、可能な生物学的作用機構を識別するためにクラスタリングおよび統計的方法などの大量のデータに対するデータマイニングを実行することができる数学的および計算技術を使用し始めた。

【0003】

以前の研究では、生物学的過程に対する1つまたは複数の攪乱の結果として生じる遺伝子発現の変化の特徴的サイン (s i g n a t u r e) を発見する可能性と、付加的なデータセット内にそのシグネチャが存在することのその後のスコア化とを調査した。この点に関する大半の研究は、疾患の表現型と相関するサインを識別し、スコア化することを伴った。これらの表現型派生サインは、著しい分類能力を備えるが、単一の特定の攪乱とサインとの間の機械的關係または因果関係を欠いている。したがって、これらのサインは、多くの場合未知の機構 (1 つまたは複数) により、同じ疾患の表現型に至るか、またはその結果生じる複数の異なる未知の攪乱を表しうる。

【0004】

生物系におけるさまざまな個別の生物学的実体の活性が、異なる生物学的機構の活性化

10

20

30

40

50

または抑制をどのように可能にするかを理解することには1つの難題が横たわっている。遺伝子などの、個別の実体が、複数の生物学的過程（例えば、炎症および細胞増殖）に関わることがあるため、遺伝子の活性を測定するだけでは、上記活性をトリガーする基礎をなす（underlying）生物学的過程を識別するには十分でない。

ランダムウォーク法は、ネットワークポロジータを特徴付けるためにネットワーク解析に使用されており、例えば、Komurovらは、データにバイアスをかけたランダムウォークを定義し、単純なランダムウォークと比較する方法について記載している（非特許文献1）。しかし、Komurovの手法では、それぞれのノードが関連データを有し、ネットワークは無向であると想定しているが、確率論的結果は提供されておらず、また利用可能な感度解析がない。加えて、因果関係ネットワークモデルを使用する場合、すべての実体（モデル内のノードとして表される）を実験的証拠とリンクすることができるとは限らない。さらに、特定の実験データが集約される場合、ネットワークは、実験によって活性化された特定の機構により不均等に攪乱される可能性がある。上記のことを考慮して、この計算生物学の分野では、生体分子ネットワークモデルにおいてハイスループットデータセットを解析するためのより進化した、より良い方法が引き続き必要とされている。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】PLOS Computational Biology、2010年8月、6(8): e1000889

【発明の概要】

【課題を解決するための手段】

【0006】

要旨

本明細書では、生物系内の実体のサブセットからの測定された活性データに基づいて1つまたは複数の攪乱（perturbation）に対する生物系の応答を定量化するためのシステム、方法、および生成物について記載する。活性データおよび生物系のネットワークモデルに基づいて中心度値（centrality value）を導出するためのシステムおよび方法が記載される。現在利用可能な技法は、微小規模で生物学的実体の活性に関与する根底の機構を識別すること（identifying）に基づくものではなく、潜在的に有害な作用物質（agent）および実験条件に応じて、これらの実体が役割を果たす種々の生物学的機構の活性化の定量的評価を行うものでもない。したがって、生物学的機構を考慮してシステム全体にわたる生物学的データを解析するための、またシステムが作用物質または環境の変化に応答するときの生物系の変化を定量化するための、システムおよび方法の改善には明確な必要性がある。

【0007】

一態様では、本明細書に記載されているシステムおよび方法は、（例えば、作用物質の曝露などの処置条件に応じて、または複数の処置条件に応じて）生物系の攪乱を定量化するためのコンピュータ化された方法、および1つまたは複数のコンピュータプロセッサを対象とする。このコンピュータ化された方法は、第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップを含むことができる。生物系は複数の生物学的実体を含み、それぞれの生物学的実体は、該生物学的実体のうちの他の少なくとも1つと相互作用する。コンピュータ化された方法はまた、第2のプロセッサで、作用物質に曝露していない生物系に対応するコントロールデータの組を受け取るステップを含むこともできる。コンピュータ化された方法はさらに、第3のプロセッサで、生物系を表す計算因果関係ネットワークモデルを提供するステップを含むことができる。この計算因果関係ネットワークモデルは、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む。エッジは、対応する第1のノードに対応する第2のノードに接続する。いくつかの実装では、エッジはノード間の因果活性化関係を表す。

【0008】

10

20

30

40

50

コンピュータ化された方法はさらに、第4のプロセッサにより、ノードのサブセットの攪乱指標を計算するステップを含むことができる。攪乱指標は、ネットワークモデルに少なくとも一部は基づいて計算される。攪乱指標は、対応するノードにおける処置データとコントロールデータとの差を表し、また対応するノードの活性が攪乱から影響を受ける程度を表す。

【0009】

コンピュータ化された方法はさらに、第5のプロセッサにより、エッジの遷移確率を計算するステップを含むことができる。エッジの遷移確率は、攪乱指標に少なくとも一部は基づいて計算することができる。エッジの遷移確率は、対応する第1のノードから対応する第2のノードへの遷移の尤度 (likelihood) を表す。このような遷移確率により、マルコフ連鎖を定義することができる。

10

【0010】

最後に、コンピュータ化された方法はさらに、第6のプロセッサにより、ノードの中心度値を生成するステップを含むことができる。ノードの中心度値は、遷移確率に少なくとも一部は基づいて生成することができ、中心度値は、ネットワークモデル内の対応するノードの相対的重要度を表す。

【0011】

いくつかの実装では、攪乱指標は、対応するノードから下流のノードの活性尺度の一次結合である。いくつかの実装では、エッジの遷移確率は、対応する第2のノードの攪乱指標に少なくとも一部は基づいている。このような実装では、エッジの遷移確率は、第2のノードの攪乱指標の一次関数とすることができる。

20

【0012】

いくつかの実装では、コンピュータ化された方法はさらに、第7のプロセッサにより、ノードを定常状態において訪問するランダムウォークの確率を表す、ノードの平衡確率を計算するステップを含む。このような実装では、第6のプロセッサは、平衡確率に少なくとも一部は基づいて中心度値を生成することができる。

【0013】

いくつかの実装では、第6のプロセッサは、他のノードへの連続する訪問の間の、対応するノードへのランダムウォークの予想される訪問の回数に少なくとも一部は基づいて、対応するノードの中心度値を生成する。このような実装では、中心度値は、ネットワーク内のノードすべてにわたって予想される訪問の回数の一次結合とすることができる。

30

【0014】

いくつかの実装では、中心度値は、攪乱指標に基づかない単純遷移確率に少なくとも一部は基づいて生成される単純中心度値によって正規化される。

【0015】

いくつかの実装では、第1から第6までのそれぞれのプロセッサは、単一のプロセッサまたは単一のコンピューティングデバイス内に収められている。他の実装では、第1から第6までのプロセッサの1つまたは複数、複数のプロセッサまたはコンピューティングデバイスにわたって分散される。

【0016】

いくつかの実装では、計算因果ネットワークモデルは、潜在的原因を表すノードと1つまたは複数の測定された量を表すノードとの間に存在する因果関係の組を含む。このような実装では、活性尺度は、倍率変化を含みうる。倍率変化は、対照データと処置データとの間、または異なる処置条件を表すデータの2つの組の間で、初期値から最終値までノード測定値がどれだけ変化するかを記述する数であるものとしてよい。倍率変化数は、これら2つの条件の間の生物学的実体の活性の倍率変化の対数を表すものとしてよい。それぞれのノードに対する活性尺度は、各ノードによって表される生物学的実体に対する処置データと対照データとの差の対数を含みうる。いくつかの実装では、コンピュータ化された方法は、プロセッサを使って、生成されたスコアのそれぞれについて信頼区間を生成するステップを含む。

40

50

【0017】

いくつかの実装では、生物系のサブセットは、限定はしないが、細胞増殖機構、細胞ストレス機構、細胞炎症機構、アポトーシス、老化、オートファジー、またはネクロプーシスの機構、およびDNA修復機構のうち少なくとも1つを含む。作用物質は、限定はしないが、生物系に存在も由来もしない分子または実体を含む異物を含むことができる。作用物質は、限定はしないが、毒素、治療化合物、刺激物、弛緩物質、天然物、製造物および食物を含むことができる。作用物質は、限定はしないが、タバコを加熱することによって発生したエアロゾル、タバコを燃焼させることによって発生したエアロゾル、タバコの煙、および紙巻きタバコの煙、のうちの少なくとも1つを含むことができる。作用物質は、限定はしないが、カドミウム、水銀、クロム、ニコチン、タバコ特有のニトロソアミン類およびその代謝物(4-(メチルニトロソアミノ(methyl nitrosoamino))-1-(3-ピリジル)-1-ブタノン(NNK)、N'-ニトロソノルニコチン(NNN)、N-ニトロソアナタピン(NAT)、N-ニトロソアナバシン(NAB)、および4-(メチルニトロソアミノ)-1-(3-ピリジル)-1-ブタノール(NNAL))を含むことができる。いくつかの実装では、作用物質は、ニコチン置換療法に使用される生成物を含む。

10

【0018】

別の態様では、本明細書に記載されているシステムおよび方法は、生物系の攪乱を定量化するためのコンピュータ化された方法、および1つまたは複数のコンピュータプロセスを対象とする。このコンピュータ化された方法は、第1のプロセッサで、第1の処置データの組を受け取るステップと、第2のプロセッサで、第2の処置データの組を受け取るステップとを含むことができる。コンピュータ化された方法はさらに、第3のプロセッサで、計算因果関係ネットワークモデルを提供するステップを含むことができる。このネットワークモデルは、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む。コンピュータ化された方法はさらに、第4のプロセッサにより、ノードのサブセットの攪乱指標を計算するステップを含むことができる。攪乱指標は、ネットワークモデルに少なくとも一部は基づいて計算することができ、対応するノードにおける第1と第2の処置データ間の差を表すことができる。コンピュータ化された方法はさらに、第5のプロセッサにより、対応するノードの中心度値を生成するステップを含むことができる。中心度値は、攪乱指標に少なくとも一部は基づいて生成することができ、ネットワークモデル内の対応するノードの相対的重要度を表す。コンピュータ化された方法はさらに、第6のプロセッサにより、第2のノードの攪乱指標に関する第1のノードの中心度値の偏導関数を計算するステップを含むことができる。この偏導関数は、ネットワークモデルのトポロジー感度尺度(topological sensitivity measure)を表す。いくつかの実装では、偏導関数を計算するステップは、第1のノードの中心度値の変化に及ぼす第2のノードの攪乱指標の変化の影響を決定するステップを含む。

20

30

【0019】

別の態様では、本明細書に記載されているシステムおよび方法は、生物系に対する攪乱の影響を視覚化するためのコンピュータ化された方法、および1つまたは複数のコンピュータプロセスを対象とする。このコンピュータ化された方法は、第1のプロセッサで、計算因果関係ネットワークモデルを提供するステップを含むことができる。このネットワークモデルは、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む。コンピュータ化された方法はさらに、第2のプロセッサにより、対応するノードの中心度値を生成するステップを含むことができる。この中心度値は、ネットワークモデルに少なくとも一部は基づいて生成することができ、ネットワークモデル内の対応するノードの相対的重要度を表すことができる。コンピュータ化された方法はさらに、第3のプロセッサにより、ネットワークモデルに対する攪乱の影響を表すためのスペクトル変換ベクトル上への中心度値の射影を計算するステップを含むことができる。いくつかの実装では、中心度値の射影を計算するステップは、中心度値をフィルタリングするステップを含む。いくつかの実装では、コンピュータ化された方法はさらに、ネットワークモデルを表

40

50

示するステップと、表示されたネットワークモデルの上に中心度値の射影の1つまたは複数の構成要素を表示するステップとを含む。いくつかの実装では、ネットワークモデル内のエッジは無向である。

【0020】

別の態様では、本明細書に記載されているシステムおよび方法は、生物系の攪乱を定量化するためのコンピュータ化された方法、および1つまたは複数のコンピュータプロセスを対象とする。このコンピュータ化された方法は、第1のプロセッサで、計算因果関係ネットワークモデルを提供するステップを含むことができる。このネットワークモデルは、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む。コンピュータ化された方法はさらに、第2のプロセッサにより、対応するノードの中心度値を生成するステップを含むことができる。この中心度値は、ネットワークモデルに少なくとも一部は基づいて生成することができ、またネットワークモデル内の対応するノードの重要性の相対的程度を表すことができる。コンピュータ化された方法はさらに、第3のプロセッサにより、中心度値を集約して、生物系の攪乱を表すネットワークモデルのスコアを生成するステップを含むことができる。いくつかの実装では、スコアはスカラー値である。いくつかの実装では、中心度値を集約するステップは、中心度値の一次結合を計算するステップを含む。いくつかの実装では、中心度値を集約するステップは、中心度値のスペクトル変換の一次結合を計算するステップを含む。

【0021】

本明細書に記載されているコンピュータ化された方法は、それぞれが1つまたは複数のプロセッサを備える1つまたは複数のコンピューティングデバイスを有するコンピュータ化されたシステムで実装されうる。一般的に、本明細書に記載されているコンピュータ化されたシステムは、コンピュータ、マイクロプロセッサ、ロジックデバイス、またはハードウェア、ファームウェア、およびソフトウェアを用いて本明細書に記載されているコンピュータ化された方法のうちの1つまたは複数を実施するように構成された他のデバイスもしくはプロセッサなどの、1つまたは複数の処理デバイスを備える、1つまたは複数のエンジンを具備することができる。いくつかの実装では、上記コンピュータ化されたシステムは、システム応答プロファイルエンジン、ネットワークモデリングエンジン、およびネットワークスコア化エンジンを備える。上記エンジンは、ときどき相互接続することができ、攪乱データベース、測定可能要素データベース、実験データデータベース、および文献データベースを含む、1つまたは複数のデータベースにさらにときどき接続されうる。本明細書に記載されているコンピュータ化されたシステムは、ネットワークインターフェースを通じて通信する1つまたは複数のプロセッサおよびエンジンを有する分散型のコンピュータ化されたシステムを含みうる。このような実装は、複数の通信システム上で分散コンピューティングを実行するのに適し得る。

例えば、本願発明は以下の項目を提供する。

(項目1)

生物系のネットワークモデル内のノードの距離を決定するコンピュータ化された方法であって、

第1のプロセッサで、作用物質に対する生物系の応答に対応する処置データの組を受け取るステップであって、該生物系が複数の生物学的実体を含み、それぞれの生物学的実体が該生物学的実体のうちの他の少なくとも1つと相互作用するステップと、

第2のプロセッサで、該作用物質に曝露していない該生物系に対応するコントロールデータの組を受け取るステップと、

第3のプロセッサで、計算因果関係ネットワークモデルを提供するステップであって、該計算因果関係ネットワークモデルが該生物系を表すと共に、

該生物学的実体を表すノード、および

該生物学的実体間の関係を表すエッジを含み、エッジが、対応する第1のノードを、対応する第2のノードに接続するステップと、

第4のプロセッサにより、該ネットワークモデルに少なくとも一部は基づいて該ノード

のサブセットの攪乱指標を計算するステップであって、攪乱指標が、対応するノードにおける該処置データと該コントロールデータの間の差を表し、また該対応するノードの活性が該攪乱から影響を受ける程度を表すステップと、

第5のプロセッサにより、該攪乱指標に少なくとも一部は基づいて該エッジの遷移確率を計算するステップであって、エッジの遷移確率が、該対応する第1のノードから該対応する第2のノードへの遷移の尤度を表すステップと、

第6のプロセッサにより、該遷移確率に少なくとも一部は基づいて該ノードの中心度値を生成するステップであって、中心度値が該ネットワークモデル内の対応するノードの相対的重要度を表すステップと

を含む、コンピュータ化された方法。

10

(項目2)

前記攪乱指標が、前記対応するノードから下流のノードの活性尺度の一次結合である、項目1に記載のコンピュータ化された方法。

(項目3)

エッジの前記遷移確率が前記第2のノードの前記攪乱指標の一次関数である、項目1または項目2に記載のコンピュータ化された方法。

(項目4)

第7のプロセッサにより、前記ノードを定常状態において訪問するランダムウォークの確率を表す、該ノードの平衡確率を計算するステップをさらに含む、前記項目のいずれかに記載のコンピュータ化された方法。

20

(項目5)

前記第6のプロセッサが、前記平衡確率に少なくとも一部は基づいて前記中心度値を生成する、前記項目のいずれかに記載のコンピュータ化された方法。

(項目6)

前記第6のプロセッサが、他のノードへの連続する訪問の間の、対応するノードへのランダムウォークの予想される訪問の回数に少なくとも一部は基づいて、該対応するノードの前記中心度値を生成する、前記項目のいずれかに記載のコンピュータ化された方法。

(項目7)

前記攪乱指標が、前記対応するノードにおける前記処置データと前記コントロールデータの間の差を表す倍率変化値にさらに基づく、前記項目のいずれかに記載のコンピュータ化された方法。

30

(項目8)

第1のプロセッサで、第1の処置データの組を受け取るステップと、
第2のプロセッサで、第2の処置データの組を受け取るステップと、
第3のプロセッサで、

生物学的実体を表すノード、および

該生物学的実体の間の関係を表すエッジを含む計算因果関係ネットワークモデルを提供するステップと、

第4のプロセッサにより、該ネットワークモデルに少なくとも一部は基づいて該ノードのサブセットの攪乱指標を計算するステップであって、攪乱指標が、対応するノードにおける該第1の処置データと該第2の処置データの間の差を表すステップと、

40

第5のプロセッサにより、該攪乱指標に少なくとも一部は基づいて、対応するノードの中心度値を生成するステップであって、中心度値が該ネットワークモデル内の該対応するノードの相対的重要度を表すステップと、

第6のプロセッサにより、第2のノードの該攪乱指標に関する第1のノードの中心度値の偏導関数を計算するステップであって、該偏導関数が該ネットワークモデルのトポロジ-感度尺度を表すステップと

を含む、コンピュータ化された方法。

(項目9)

前記偏導関数を計算するステップが、前記第1のノードの前記中心度値の変化に及ぼす

50

前記第2のノードの前記攪乱指標の変化の影響を決定するステップを含む、項目8に記載のコンピュータ化された方法。

(項目10)

第1のプロセッサで、

生物学的実体を表すノード、および

該生物学的実体間の関係を表すエッジを含む計算ネットワークモデルを提供するステップと、

第2のプロセッサにより、該ネットワークモデルに少なくとも一部に基づいて、対応するノードの中心度値を生成するステップであって、中心度値が該ネットワークモデル内の該対応するノードの相対的重要度を表すステップと、

第3のプロセッサにより、該ネットワークモデルに対する攪乱の影響を表すためのスペクトル変換ベクトル上への該中心度値の射影を計算するステップと
を含む、コンピュータ化された方法。

(項目11)

前記中心度値の射影を計算するステップが、前記中心度値をフィルタリングするステップを含む、項目10に記載のコンピュータ化された方法。

(項目12)

生物系の攪乱を定量化するコンピュータ化された方法であって、

第1のプロセッサで、

生物学的実体を表すノード、および

該生物学的実体間の関係を表すエッジを含む計算因果関係ネットワークモデルを提供するステップと、

第2のプロセッサにより、該ネットワークモデルに少なくとも一部に基づいて対応するノードの中心度値を生成するステップであって、中心度値が該ネットワークモデル内の該対応するノードの相対的重要度を表すステップと、

第3のプロセッサにより、該中心度値を集約して、該生物系の攪乱を表す該ネットワークモデルのスコアを生成するステップと
を含む、コンピュータ化された方法。

(項目13)

前記スコアがスカラー値である、項目12に記載のコンピュータ化された方法。

(項目14)

前記中心度値を集約するステップが、該中心度値の一次結合を計算するステップを含む、項目12または13に記載のコンピュータ化された方法。

(項目15)

前記中心度値を集約するステップが、該中心度値のスペクトル変換の一次結合を計算するステップを含む、項目12または13に記載のコンピュータ化された方法。

【0022】

本開示のさらなる特徴、その特質、およびさまざまな利点は、図面全体を通して類似の参照文字が類似の部品を指す付属の図面と併せて、以下の詳細な記載を考慮に入れることで、明らかになる。

【図面の簡単な説明】

【0023】

【図1】 図1は、攪乱への生物学的ネットワークの応答を定量化するための例示的なコンピュータ化されたシステムのブロック図である。

【0024】

【図2】 図2は、ネットワーク攪乱振幅(NPA)スコアを計算することによって攪乱への生物学的ネットワークの応答を定量化するための例示的なプロセスの流れ図である。

【0025】

【図3】 図3は、2種類の作用物質、2つのパラメータ、およびN個の生物学的実体に対するデータを含むシステム応答プロファイルの基礎となるデータの図形表現である。

10

20

30

40

50

【 0 0 2 6 】

【図 4 A】図 4 A および図 4 B は、いくつかの生物学的実体を有する生物学的ネットワークおよびそれらの関係の計算モデルを示す図である。

【図 4 B】図 4 A および図 4 B は、いくつかの生物学的実体を有する生物学的ネットワークおよびそれらの関係の計算モデルを示す図である。

【 0 0 2 7 】

【図 5】図 5 は、生物学的ネットワーク内のノードの中心度値を生成する例示的プロセスを示す流れ図である。

【 0 0 2 8 】

【図 6】図 6 は、1 組のノードの攪乱指標を生成する例示的プロセスを示す、図 5 の一部分のより詳細な流れ図である。

10

【 0 0 2 9 】

【図 7】図 7 は、ネットワークについて強化ランダムウォークを定義する例示的プロセスを示す、図 5 の一部分のより詳細な流れ図である。

【 0 0 3 0 】

【図 8】図 8 は、1 組のノードの中心度値を計算する例示的プロセスを示す、図 5 の一部分のより詳細な流れ図である。

【 0 0 3 1 】

【図 9】図 9 は、生物学的攪乱の影響を定量化する例示的な分散型のコンピュータ化されたシステムのブロック図である。

20

【 0 0 3 2 】

【図 1 0】図 1 0 は、本明細書に記載されているコンピュータ化されたシステムのいずれかにおけるコンポーネントのうちのいずれかを実装するために使用されうる例示的なコンピュータ化されたシステムを示すブロック図である。

【 0 0 3 3 】

【図 1 1】図 1 1 は、因果関係ネットワークモデルの簡略図である。

【 0 0 3 4 】

【図 1 2】図 1 2 は、因果関係ネットワークの簡略図である。

【 0 0 3 5 】

【図 1 3】図 1 3 は、ネットワーク内の中心度値の射影スペクトル成分の簡略図である。

30

【図 1 4】図 1 4 は、ネットワーク内の中心度値の射影スペクトル成分の簡略図である。

【 0 0 3 6 】

【図 1 5 - 1】図 1 5 は、細胞増殖に関して肺に注目した因果関係ネットワークの一例の図である。

【図 1 5 - 2】図 1 5 は、細胞増殖に関して肺に注目した因果関係ネットワークの一例の図である。

【図 1 5 - 3】図 1 5 は、細胞増殖に関して肺に注目した因果関係ネットワークの一例の図である。

【図 1 5 - 4】図 1 5 は、細胞増殖に関して肺に注目した因果関係ネットワークの一例の図である。

40

【 0 0 3 7 】

【図 1 6】図 1 6 は、ノード細胞増殖の中心度値に関する実験結果のグラフである。

【発明を実施するための形態】

【 0 0 3 8 】

詳細な説明

本願の範囲内で使用される技術用語および表現には一般に、関連技術において通常適用される意味が与えられる。「含む (comprising)」という語は、他の要素またはステップを除外せず、不定冠詞「a」または「an」は複数を除外しない。特に属性または値に関連した「本質的に (essentially)」、「約 (about)」、「およそ (approximately)」などの語はまた、それぞれその属性を厳密に定

50

義し、またはその値を厳密に定義する。本明細書では、生物系が作用物質によって攪乱された場合の生物系内の変化の大きさを定量的に評価する計算システム、コンピュータ化された方法および生成物について記載する。いくつかの実装は、生物系の一部内の変化の大きさを表現する数値を計算するための方法を含む。この計算では、入力として、作用物質によって生物系が攪乱される制御された実験の組から得られたデータの組を使用する。次いで、データが、生物系の特徴のネットワークモデルに適用される。ネットワークモデルは、シミュレーションおよび分析のための基盤 (s u b s t r a t e) として使用され、生物系内の目的の特徴を使用可能にする生物学的機構および経路を表す。この機構および経路の特徴または一部は、生物系の疾病および有害作用の病理に関与しうる。通常状態下および作用物質による攪乱下を含む、さまざまな条件の下での多数の生物学的実体のステータスに関するデータによって占められるネットワークモデルを構築するために、データベースで表されている生物系の従来知識が使用される。使用されるネットワークモデルは、それが攪乱にตอบสนองするさまざまな生物学的実体のステータスの変化を表し、生物系に対する作用物質の影響の定量的および客観的評価を得ることができるという点で、動的である。これらの計算方法を運用するためのコンピュータシステムおよび生成物も提供される。

10

【 0 0 3 9 】

本開示のコンピュータ化された方法によって生成された数値は、とりわけ、1つまたは複数の製造された産物（安全性評価または比較のため）、栄養補強物を含む治療化合物（効力または健康上の利益の判定のため）、および環境作用物質（長期曝露の危険性ならびに有害作用および発症との関係の予測のため）によって引き起こされる望ましい、または有害な生物学的作用の大きさを判定するために使用されうる。

20

【 0 0 4 0 】

一態様では、本明細書に記載されているシステムおよび方法は、攪乱された生物学的機構のネットワークモデルに基づき、攪乱された生物系の変化の大きさを表す計算された数値を提供する。本明細書でネットワーク攪乱振幅 (N P A) スコアと称される数値は、定義されている生物学的機構におけるさまざまな実体のステータスの変化の概略を表すために使用されうる。異なる作用物質または異なる種類の攪乱に対して得られた数値は、生物系の特徴としてそれ自体を使用可能にするか、またはそれ自体を現す生物学的機構に対する異なる作用物質または攪乱の影響を相対比較するために使用することができる。そこで、N P A スコアは、異なる攪乱に対する生物学的機構の応答を測定するために使用することができる。「スコア」という用語は、本明細書では、生物系における変化の大きさの量的尺度を与える値または値の組を一般的に指す。このようなスコアは、サンプルまたは被験体から得られた1つまたは複数のデータセットを使用して、当技術分野で公知の、本明細書で開示されている方法による、さまざまな数学的アルゴリズムおよび計算アルゴリズムのうちのいずれかを使用して計算される。

30

【 0 0 4 1 】

N P A スコアは、研究者および臨床医による診断、実験計画、治療決定、およびリスクアセスメントの改善を助けることができる。例えば、N P A スコアは、毒物学的分析において候補となる生物学的機構の組をスクリーニングして、潜在的に有害な作用物質への曝露で最も影響を受けそうなものを識別するために使用することができる。攪乱へのネットワークの応答の尺度を提供することによって、これらのN P A スコアは、細胞レベル、組織レベル、器官レベル、または生物レベルで出現する表現型または生物学的転帰との分子事象の相関（実験データによって測定されている場合）を可能にすることができる。臨床医は、N P A 値を使用して、作用物質によって影響される生物学的機構を患者の生理学的状態と比較し、作用物質に曝露されたときに患者がどのような健康上の危険性または利益を受ける可能性が最も高いかを判定することができる（例えば、免疫無防備状態の (i m m u n o - c o m p r o m i s e d) 患者は、強い免疫抑制応答を引き起こす作用物質に対して特に脆弱であり得る）。

40

【 0 0 4 2 】

50

図1は、攪乱へのネットワークモデルの応答を定量化するためのコンピュータ化されたシステム100のブロック図である。特に、システム100は、システム応答プロファイルエンジン110、ネットワークモデリングエンジン112、およびネットワークスコア化エンジン114を備える。エンジン110、112、および114は、ときどき相互接続され、攪乱データベース102、測定可能要素データベース104、実験データデータベース106、および文献データベース108を含む、1つまたは複数のデータベースにときどきさらに接続される。本明細書で使用されているように、エンジンは、コンピュータ、マイクロプロセッサ、ロジックデバイス、またはハードウェア、ファームウェア、およびソフトウェアを用いて1つまたは複数の計算オペレーションを実行するように構成された、図10を参照しつつ記載されているような他の1つまたは複数のデバイスなどの、1つまたは複数の処理デバイスを備える。

10

【0043】

図2は、一実装による、ネットワーク攪乱振幅(NPA)スコアを計算することによって攪乱への生物学的ネットワークの応答を定量化するためのプロセス200の流れ図である。プロセス200のステップは、図1のシステム100のさまざまなコンポーネントによって実行されるように記述されるが、これらのステップはいずれも、ローカルもしくはリモートの任意の好適なハードウェアコンポーネントまたはソフトウェアコンポーネントによって実行することができ、また任意の適切な順序に配置構成されるか、または並列実行されうる。ステップ210で、システム応答プロファイル(SRP)エンジン110は、さまざまな異なるソースから生物学的データを受け取り、データそれ自体は、さまざまな異なる型のものであってよい。データは、生物系が攪乱される実験からのデータ、さらには対照データを含む。ステップ212で、SRPエンジン110は、生物系内の1つまたは複数の実体が、生物系に対する作用物質の提示に反応して変化する程度の表現である、システム応答プロファイル(SRP)を生成する。ステップ214で、ネットワークモデリングエンジン112は、その1つが作用物質または目的の特徴に関連するものとして選択される複数のネットワークモデルを含む1つまたは複数のデータベースを提供する。この選択は、系の生物学的機能の基礎をなす機構の従来からの知識に基づいてなされうる。いくつかの実装では、ネットワークモデリングエンジン112は、システム応答プロファイルを用いるシステム内の実体、データベース内のネットワーク、および文献にすでに記載されているネットワークの間の因果関係を抽出し、それにより、ネットワークモデルの生成、精密化、または拡張を行うことができる。ステップ216で、ネットワークスコア化エンジン114は、ネットワークモデリングエンジン112によってステップ214で識別されたネットワークおよびSRPエンジン110によってステップ212で生成されたSRPを使用してそれぞれの攪乱についてNPAスコアを生成する。NPAスコアは、生物学的実体(ネットワークによって表される)の間の基礎をなす関係の状況において攪乱または処置(SRPで表される)への生物学的応答を定量化する。

20

30

【0044】

本開示との関連での生物系は、機能的部分を含む、生物または生物の一部を含み、該生物は本明細書では被験体と称される。上記被験体は、一般的に、ヒトを含む、哺乳類である。上記被験体は、ヒト集団における個別のヒトとすることができる。本明細書で使用されているような「哺乳類」という用語は、限定はしないが、ヒト、ヒト以外の霊長類、マウス、ラット、イヌ、ネコ、ウシ、ヒツジ、ウマ、およびブタを含む。ヒト以外の哺乳類は、有利には、ヒトの疾患のモデルを提供するために使用されうる被験体として使用されうる。ヒト以外の被験体は、非改変であるか、または遺伝子組み換え動物(例えば、トランスジェニック動物、または1つもしくは複数の遺伝子変異またはサイレンシングされた遺伝子(1つまたは複数)を持つ動物)とすることができる。上記被験体は、オスまたはメスとすることができる。上記操作の目的に応じて、上記被験体は、目的の作用物質に曝露させた被験体とすることができる。上記被験体は、必要に応じて研究までの時間を含む、長期間にわたって作用物質に曝露させた被験体とすることができる。上記被験体は、一定の期間にわたって作用物質に曝露させたか、または該作用物質ともはや接触していない

40

50

被験体とすることができる。上記被験体は、疾患を有しているものとして診断または識別された被験体とすることができる。上記被験体は、疾患または有害な健康状態の処置をすでに受けたか、または今受けている最中である被験体とすることができる。上記被験体は、特定の健康状態または疾患に対する1つまたは複数の症状もしくは危険因子を示す被験体とすることもできる。上記被験体は、疾患にかかりやすい被験体とすることができ、症状を示すか、または無症候性であるかのいずれかであってよい。いくつかの実装では、目的の疾患または健康状態は、作用物質への曝露、または長期間にわたる作用物質の使用に関連する。いくつかの実装によれば、上記システム100(図1)は、攪乱の種類または目的とする転帰に関連する1つまたは複数の生物系およびその機能の機構(まとめて、「生物学的ネットワーク」または「ネットワークモデル」)のコンピュータ化されたモデルを含むか、またはそれを生成する。

10

【0045】

上記操作の環境(context)に応じて、生物系は、それが、集団における個別の生物、一般的に生物、器官、組織、細胞型、細胞小器官、細胞成分、または特定の個人の細胞(1つまたは複数)の機能に関係するとおりに異なるレベルで定義されうる。それぞれの生物系は、1つまたは複数の生物学的機構または経路を備え、上記操作はその系の機能的特徴として現れる。ヒト健康状態の定義された特徴を再現し、目的の作用物質への曝露について適している動物系は、好ましい生物系である。疾患の原因または病理に関わる細胞型および組織を反映する細胞および器官型系も、好ましい生物系である。in vivoでヒト生物学をできる限り反復する初代細胞または器官培養物を優先することも可能である。また、in vitroのヒト細胞培養物と動物モデルからin vivoで導出される最も等価の培養物とをマッチさせることも重要である。これは、基準系としてin vitroでマッチした系を使用してin vivoの動物モデルからヒト生物学への翻訳連続体(translational continuum)の創製を可能にする。したがって、本明細書に記載されているシステムおよび方法とともに使用することが企図されている生物系は、限定はしないが、機能的特徴(例えば、生物学的機能、生理学的機能、または細胞機能)、小器官、細胞型、組織種類、器官、発達段階、または上記の組み合わせによって定義されうる。生物系の例として、限定はしないが、肺系、外皮系、骨格系、筋肉系、神経系(例えば、中枢神経および末梢神経)、内分泌系、心血管系、免疫系、循環系、呼吸器系、泌尿器系、腎臓系、胃腸系、結腸直腸系、肝臓系、および生殖器系が挙げられる。生物系の他の例として、限定はしないが、上皮細胞、神経細胞、血液細胞、結合組織細胞、平滑筋細胞、骨格筋細胞、脂肪細胞、卵細胞、精子細胞、幹細胞、肺細胞、脳細胞、心臓細胞、喉頭細胞、咽頭細胞、食道細胞、胃細胞、腎細胞、肝細胞、乳腺細胞、前立腺細胞、膵臓細胞、島細胞、精巣細胞、膀胱細胞、頸部細胞、子宮細胞、結腸細胞、および直腸細胞のさまざまな細胞機能が挙げられる。これらの細胞のうちのいくつかは、in vitroで培養されるか、または適切な培養条件の下で無期限にin vitroで維持される細胞系の細胞であるものとしてよい。細胞機能の例として、限定はしないが、細胞増殖(例えば、細胞分裂)、変性、再生、老化、核による細胞活性の制御、細胞間シグナル伝達、細胞分化、細胞脱分化、分泌、遊走、食作用、修復、アポトーシス、および発生プログラミングが挙げられる。生物系として考えることができる細胞成分の例として、限定はしないが、細胞質、細胞骨格、膜、リボソーム、ミトコンドリア、核、小胞体(ER)、ゴルジ体、リソソーム、DNA、RNA、タンパク質、ペプチド、および抗体が挙げられる。

20

30

40

【0046】

生物系における攪乱は、該生物系の1つまたは複数の部分を曝露させるか、または接触させることを通じて一定期間にわたって1つまたは複数の作用物質によって引き起こされうる。作用物質は、すべての構成成分が識別や特徴付けがなされるとは限らない混合物を含む、単一の物質または物質の混合物もしくは複数の(例えば、1つまたは複数)物質とすることができる。作用物質またはその構成成分の化学的および物理的特性は完全に特徴付けられない場合もある。作用物質は、その構造、その構成成分、またはある条件の下で

50

該作用物質を生成する供給源によって定義されうる。作用物質の一例は、上記生物系中に存在も由来もしない分子もしくは実体であり、該生物系と接触した後にその作用物質から生成される任意の中間体または代謝産物である異物である。作用物質は、炭水化物、タンパク質、脂質、核酸、アルカロイド、ビタミン、金属、重金属、ミネラル、酸素、イオン、酵素、ホルモン、神経伝達物質、無機化合物、有機化合物、環境作用物質、微生物、粒子、環境条件、環境的影響力、または物理的力のうちの1つまたは複数であってよい。作用物質の非限定的な例として、限定はしないが、栄養素、代謝廃棄物、毒物、麻薬、毒素、治療化合物、刺激物質、弛緩物質、天然物、製造物、食物、病原体（プリオン、ウイルス、細菌、真菌、原生生物）、寸法がマイクロメートル範囲またはそれ未満の粒子もしくは実体、上記のものの副産物、および上記のものの混合物が挙げられる。物理的作用物質の非限定的な例として、放射線、電磁波（太陽光を含む）、温度の上昇もしくは低下、剪断力、流体圧力、放電（1つまたは複数）またはそのシーケンス、あるいは外傷が挙げられる。

10

【0047】

少なくともいくつかの、またはすべての作用物質は、閾値濃度で存在していない限り、または一定期間生物系と接触していない限り、またはその両方の組み合わせが生じていない限り生物系を攪乱しえない。攪乱を結果として引き起こす作用物質への曝露または接触は、用量に関して定量化されうる。したがって、攪乱は、作用物質（1つまたは複数）への長期的曝露の結果生じうる。曝露の期間は、時間の単位で、曝露の頻度で、または上記被験体の実際のもしくは推定される寿命における時間のパーセンテージで表すことができる。攪乱は、上記生物系の1つまたは複数の部分に、作用物質の供給源から作用物質（上に記載されているような）を供給しないようにするか、または作用物質の供給を制限することによって引き起こされることもある。例えば、攪乱は、1つまたは複数の栄養素、水、炭水化物類、タンパク質、脂質、アルカロイド、ビタミン、ミネラル、酸素、イオン、酵素、ホルモン、神経伝達物質、抗体、サイトカイン、光の供給不足もしくは欠如によって、または生物のいくつかの部分の移動を制約することによって、または運動を抑圧もしくは要求することによって引き起こされうる。それらの組み合わせが企図される。

20

【0048】

少なくともいくつかの、またはすべての作用物質は、上記生物系のどの部分（1つまたは複数）が曝露されるか、および曝露条件によって異なる攪乱を引き起こしうる。作用物質の非限定的な例は、タバコを加熱することによって発生したエアロゾル、タバコを燃焼させることによって発生したエアロゾル、タバコの煙、紙巻きタバコの煙、およびこれらのガス状構成成分または粒子状構成成分のいずれかを含みうる。作用物質のさらなる非限定的な例として、カドミウム、水銀、クロム、ニコチン、タバコ特有のニトロソアミン類およびその代謝物（4-（メチルニトロソアミノ）-1-（3-ピリジル）-1-ブタノン（NNK）、N'-ニトロソノルニコチン（NNN）、N-ニトロソアナタピン（NAT）、N-ニトロソアナバシン（NAB）、4-（メチルニトロソアミノ）-1-（3-ピリジル）-1-ブタノール（NNAL）など）、およびニコチン置換療法のために使用される生成物が挙げられる。作用物質または複合刺激物についての曝露処方計画は、毎日の設定における曝露の範囲および環境を反映すべきである。一群の標準的な曝露処方計画は、同様に定義の明確な（*equally well-defined*）実験系に体系的に適用されるように設計されうる。それぞれのアッセイは、初期と後期の事象の両方を捕らえ、代表的な用量範囲が確実にカバーされるように時間および用量依存のデータを収集するように設計することが可能である。しかし、当業者であれば、本明細書に記載されているシステムおよび方法が取り扱われる適用に適しているように適合され改変されうること、また本明細書において設計されているシステムおよび方法が他の好適な適用において使用されうること、またそのような他の追加および改変が本発明の範囲から逸脱しないことを理解する。

30

40

【0049】

さまざまな実装において、遺伝子の発現、タンパク質の発現もしくはタンパク質の代謝

50

回転、マイクロRNAの発現もしくはマイクロRNAの代謝回転、翻訳後修飾、タンパク質修飾、転座、抗体産生代謝産物プロファイル、または上記のうちの2つ以上のものの組み合わせに対する系全体にわたるハイスループット測定が、各対照を含むさまざまな条件の下で生成される。これらは、一般的に、上記評価のためのアンカーとして働き、疾患の原因における明確なステップを表すことができるので、機能転帰測定は、本明細書に記載されている方法において望ましい。

【0050】

本明細書で使用されているような「サンプル」は、被験体または実験系（例えば、細胞、組織、器官、または動物全体）から分離される任意の生体サンプルを指す。サンプルは、限定はしないが、単細胞もしくは多細胞、細胞画分、組織生検、切除組織、組織抽出物、組織、組織培養抽出物、組織培基、吐き出されたガス、全血、血小板、血清、血漿、赤血球、白血球、リンパ球、好中球、マクロファージ、B細胞もしくはそのサブセット、T細胞もしくはそのサブセット、造血細胞のサブセット、内皮細胞、滑液、リンパ液、腹水、間質液、骨髄、脳脊髄液、胸水、腫瘍浸潤物、唾液、粘液、痰、精液、汗、尿、または任意の他の体液を含むことができる。サンプルは、限定はしないが、静脈穿刺、排泄、生検、針吸引、洗浄、擦過、外科的切除、または当技術分野で公知の他の手段を含む手段によって被験体から得ることができる。

10

【0051】

操作中に、所与の生物学的機構、転帰、攪乱、または上記の組み合わせについて、上記システム100は、処置条件にตอบสนองしてネットワークにおける生物学的実体のステータスの変化の定量的尺度である、ネットワーク攪乱振幅（NPA）値を生成することができる。

20

【0052】

上記システム100（図1）は、目的の健康状態、疾患、または生物学的転帰に関連する1つまたは複数のコンピュータ化されたネットワークモデル（1つまたは複数）を備える。これらのネットワークモデルのうちの1つまたは複数は、以前の生物学的知識に基づいており、外部ソースからアップロードされ、該システム100内で精選されうる。上記モデルは、測定結果に基づき上記システム100内で新たに生成することもできる。測定可能な要素は、以前の知識を用いることで生物学的ネットワークモデルへと因果的に組み込まれる。以下では、ネットワークモデルを生成もしくは精密化するために使用されうる目的の生物系における変化を表す、または攪乱へのตอบสนองを表すデータの型について記載する。

30

【0053】

図2を再び参照すると、ステップ210で、上記システム応答プロファイル（SRP）エンジン110は、生物学的データを受け取る。上記SRPエンジン110は、さまざまな異なるソースからこのデータを受け取ることができ、該データそれ自体は、さまざまな異なる型のものであり得る。上記SRPエンジン110によって使用される生物学的データは、文献、データベース（医薬品または医療デバイスの前臨床試験、臨床試験、および臨床後試験からのデータを含む）、ゲノムデータベース（ゲノム配列および発現データ、例えば、National Center for Biotechnology InformationによるGene Expression OmnibusまたはEuropean Bioinformatics InstituteによるArrayExpress（Parkinsonら、2010年、Nucl. Acids Res.、doi: 10.1093/nar/gkq1040. Pubmed ID 21071405）、市販のデータベース（例えば、Gaithersburg, MD, USAのGene Logic）、または実験研究から取り出すことができる。上記データは、特定の処置条件の効果、または特定の作用物質への曝露の効果の研究するように特に設計されている1つまたは複数の種を用いてin vitro実験、ex vivo実験、またはin vivo実験などの1つまたは複数の異なるソースからの生データを含むものとしてよい。in vitro実験系は、ヒトの疾患の重要側面を表す組織培養または

40

50

器官型培養（三次元培養）を含みうる。このような実装では、これらの実験のための作用物質の用量決定および曝露の処方計画は、通常の使用もしくは活性状態の間、または特別な使用もしくは活性状態の間、ヒトについて予想されうる曝露の範囲および環境を実質的に反映することができる。実験パラメータおよび試験条件は、上記作用物質および上記曝露条件の性質、問題の生物系の分子および経路、関与する細胞型および組織、目的の転帰、および疾患の原因の側面を反映することが望まれているとおりを選択されうる。特定の動物モデル由来分子、細胞、または組織は、特定のヒト分子、細胞または組織培養物とマッチさせて、動物ベースの所見の翻訳性（*translatability*）を改善することができる。

【0054】

ハイスループットの実験技術によって多くが生成されるSRPエンジン110によって受け取られるデータは、限定はしないが、核酸に関係するもの（例えば、特定DNAもしくはRNA種の絶対的または相対的な量、DNA配列、RNA配列の変化、三次構造の変化、または、配列決定によって決定されるようなメチル化パターン、特にマイクロアレイ上の核酸に対するハイブリダイゼーション、定量的ポリメラーゼ連鎖反応、あるいは当技術分野で公知の他の技術）、タンパク質/ペプチド（例えば、絶対的または相対的な量のタンパク質、タンパク質の特定の断片、ペプチド、二次または三次構造の変化、または当技術分野で公知の方法によって決定されるような翻訳後修飾）、および機能的活性（例えば、酵素活性、タンパク質分解活性、転写調節活性、輸送活性、いくつかの結合パートナーへの結合親和力）を、いくつかの条件の下で、とりわけ含む。タンパク質またはペプチドの翻訳後修飾を含む修飾は、限定はしないが、メチル化、アセチル化、ファルネシル化、ビオチン化、ステアロイル化、ホルミル化、ミリストイル化、パルミトイル化、ゲラニルゲラニル化、ペグ化、リン酸化、硫酸化、グリコシル化、糖修飾、脂質化、脂質修飾、ユビキチン化、スモイル化、ジスルフィド結合、システイニル化、酸化、グルタチオン化、カルボキシル化、グルクロン酸化、および脱アミドを含むことができる。それに加えて、タンパク質は、アマドリ反応、シッフ塩基反応、および糖化タンパク質生成物を生じるメイラード反応などの一連の反応によって翻訳後修飾されうる。

【0055】

上記データは、限定はしないが、細胞レベルでは細胞増殖、発生的運命、および細胞死を含むもの、生理学的レベルでは、肺気量、血圧、運動熟達度などの、測定された機能的転帰も含みうる。上記データは、限定はしないが、腫瘍転移、腫瘍寛解、機能消失、および疾患の特定の段階における平均余命などの、疾患活性または疾患重症の尺度も含みうる。疾患活性は、臨床的評価によって測定することができ、その結果は、値であるか、または定められた条件の下での1体または複数体の被験体からサンプル（またはサンプルの集団）の評価から得ることができる値の組である。臨床的評価は、被験体による面接またはアンケートに対する回答に基づくものとするともできる。

【0056】

このデータは、システム応答プロファイルを決定する際に使用するため明示的に生成されている場合があるか、または以前の実験でもたらされたか、または文献に公開されている場合もある。一般的に、上記データは、分子、生物学的構造、生理学的状態、遺伝形質、または表現型に関係する情報を含む。いくつかの実装では、上記データは、分子の状態、配置、量、活性、または下部構造、生物学的構造、生理学的状態、遺伝形質、または表現型の記述を含む。後に記載するように、臨床現場では、上記データは、作用物質に曝露された、ヒト被験体から得られたサンプルに対して実施されたアッセイまたはヒト被験体に関する観察結果から得られる生データまたは処理済みデータを含みうる。

【0057】

ステップ212で、システム応答プロファイル（SRP）エンジン110は、ステップ212で受け取った生物学的データに基づいてシステム応答プロファイル（SRP）を生成する。このステップは、バックグラウンド補正、正規化、倍率変化計算、有意性決定、および差次的応答の識別（例えば、差次的に発現する遺伝子）のうちの1つまたは複数

10

20

30

40

50

含みうる。SRPは、生物系内の1つまたは複数の測定された実体（例えば、分子、核酸、ペプチド、タンパク質、細胞など）が生物系に適用される攪乱（例えば、作用物質への曝露）に応答して個別に変化する程度を表す表現である。一例において、SRPを生成するために、SRPエンジン110は、所与の実験系（「システム-処置」ペア）に適用されるパラメータの所与の組（例えば、処置もしくは攪乱条件）に対する測定値の組を収集する。図3は、2つのSRP、つまり、種々のパラメータ（例えば、第1の処置作用物質への曝露の用量および時間）を用いて第1の処置306を受けるN個の異なる生物学的実体についての生物活性データを含むSRP302、および第2の処置308を受けるN個の異なる生物学的実体についての生物活性データを含む類似のSRP304を示している。SRPに含まれるデータは、生の実験データ、処理済み実験データ（例えば、外れ値を除外するためにフィルタリングされている、信頼度推定でマークされている、多数の試行にわたって平均がとられている）、計算生物学的モデルによって生成されたデータ、または科学文献から取ったデータであってよい。SRPは、絶対値、絶対変化、倍率変化、対数変化、関数、および表などの、さまざまな方法でデータを表すことができる。SRPエンジン110は、SRPをネットワークモデリングエンジン112に渡す。

【0058】

前のステップで導出されたSRPはネットワーク攪乱の大きさが、それによって決定されることになる実験データを表しているが、計算および解析のための基盤であるのは生物学的ネットワークモデルである。この解析は、生物系の特徴に関連する機構および経路の詳細なネットワークモデルの開発を必要とする。このようなフレームワークは、より古典的な遺伝子発現の解析で使用されている遺伝子リストの調査を超える機構的理解の1つの層を提供する。生物系のネットワークモデルは、動的生物系を表し、生物系のさまざまな基本的特性に関する定量的情報をアセンブルすることによって構築される数学的構築体である。

【0059】

このようなネットワークの構築は、反復プロセスである。ネットワークの境界の線引きは、目的の過程（例えば、肺における細胞増殖）に関連する機構および経路の文献調査によって導かれる。これらの経路を記述する因果関係は、ネットワークの核をなす従来の知識から抽出される。文献ベースのネットワークは、関連する表現型端点を含むハイスループットデータセットを用いて検証されうる。SRPエンジン110はデータセットを解析するために使用することができ、この結果を使用してネットワークモデルを確認し、精密化し、または生成することができる。

【0060】

図2を再び参照すると、ステップ214で、ネットワークモデリングエンジン112は、目的の生物系の特徴の基礎をなす機構（1つまたは複数）または経路（1つまたは複数）に基づくネットワークモデルとともにSRPエンジン110からのシステム応答プロファイルを使用している。いくつかの態様では、上記ネットワークモデリングエンジン112は、SRPに基づいてすでに生成されているネットワークを識別するために使用される。上記ネットワークモデリングエンジン112は、モデルへの更新および変更を受け取るためのコンポーネントを備えることができる。上記ネットワークモデリングエンジン112は、新規データを組み込み、追加の、または精密化されたネットワークモデルを生成して、ネットワーク生成のプロセスを反復することもできる。また上記ネットワークモデリングエンジン112は、1つもしくは複数のデータセットのマージまたは1つもしくは複数のネットワークのマージを円滑にすることもできる。データベースから取り出されたネットワークの組は、追加のノード、エッジ、または全く新しいネットワークによって（例えば、特定の生物学的実体によって直接調節される追加の遺伝子の記述について文献のテキストをマイニングすることによって）手動で補うことができる。これらのネットワークは、プロセススコア化を使用可能にすることができる特徴を含む。ネットワークトポロジーが維持され、因果関係のネットワークは、ネットワークにおける任意の地点から測定可能な実体まで追跡されうる。さらに、これらのモデルは動的であり、それらのモデルを組

10

20

30

40

50

み立てる (build) ために使用される仮定は、改変または言い換えることができ、異なる組織の環境および種に適合性を与えることができる。これは、新しい知識が利用可能になると反復試験および改善を可能にする。上記ネットワークモデリングエンジン 112 は、信頼度の低い、または科学文献に記載されている実験結果と食い違う対象となるノードまたはエッジを除去することができる。上記ネットワークモデリングエンジン 112 は、監督された学習または監督のない学習の方法 (例えば、計量学習、行列補完、パターン認識) を用いて推論できる追加のノードまたはエッジを備えることもできる。

【0061】

いくつかの態様において、生物系は、頂点 (またはノード) と該ノード同士を接続するエッジからなる数学的なグラフとしてモデル化される。例えば、図 4 A および図 4 B は、単純ネットワーク 400 a および 400 b をそれぞれ示す。単純ネットワーク 400 a は、9 個のノード (ノード 402 および 404 を含む) とエッジ (406 および 408) とを備える。上記ノードは、限定はしないが、化合物、DNA、RNA、タンパク質、ペプチド、抗体、細胞、組織、および器官などの、生物系における生物学的実体を表すものとしてすることができる。上記エッジは、上記ノード間の関係を表しうる。上記グラフ内のエッジは、上記ノード間の関係を表しうる。例えば、エッジは、「に結合する」関係、「で表される」関係、「発現プロファイリングに基づいて共調節される」関係、「阻害する」関係、「原稿中に共出現する」関係、または「構造要素を共有する」関係を表しうる。一般的に、これらの種類の関係は、一对のノードの間の関係を記述する。上記グラフにおけるノードは、ノード間の関係も表しうる。したがって、上記グラフで表される、関係の間の関係 (relationships between relationships)、または 1 つの関係と別の種類の生物学的実体との間の関係を表すことが可能である。例えば、化学物質を表す 2 つのノードの間の関係は、反応を表すものとしてすることができる。この反応は、反応と反応を阻害する化学物質との間の関係のノードとしてすることができる。

【0062】

グラフのエッジは、一方の頂点から別の頂点へ有向であってもよい。例えば、生物学的な文脈において、転写調節ネットワークおよび代謝ネットワークは、有向グラフとしてモデル化されうる。転写調節ネットワークのグラフモデルでは、ノードは遺伝子を表し、エッジがそれらの遺伝子間の遺伝子転写の調節関係を示す。別の例として、タンパク質間相互作用ネットワークは、生物のプロテオーム内のタンパク質間の直接的な物理的相互作用を記述し、そのようなネットワーク内の相互作用と関連付けられている方向がないことが多い。そのため、これらのネットワークは、無向エッジとしてモデル化して、エッジと関連付けられた 2 つの頂点の間に相違がないことを意味することができる。いくつかのネットワークは、有向と無向の両方のエッジを有することができる。グラフを構成する実体および関係 (つまり、ノードおよびエッジ) は、システム 100 内のデータベース内の相互に関連付けられているノードのウェブとして格納されうる。

【0063】

上記データベース内で表される知識は、さまざまな異なるソースから引き出される、さまざまな異なる種類の知識であってよい。例えば、特定のデータは、遺伝子に関する情報、および遺伝子間の関係を含む、ゲノムデータベースを表しうる。このような一例では、ノードは、がん遺伝子を表し、そのがん遺伝子ノードに接続されている別のノードは、該がん遺伝子を阻害する遺伝子を表すことができる。上記データは、タンパク質、およびタンパク質間の関係、疾患およびそれらの相互関係、ならびにさまざまな疾患状態を表すことができる。図形表現で組み合わせることができる多くの異なる型のデータが存在する。計算モデルは、例えば、DNA データセット、RNA データセット、タンパク質データセット、抗体データセット、細胞データセット、組織データセット、器官データセット、医療データセット、疫学データセット、化学データセット、毒物学データセット、患者データセット、および集団データセットにおける知識を表すノード間の関係のウェブを表すものとしてよい。本明細書で使用される場合、データセットは、定められた条件の下でサンプル (またはサンプルの群) の評価の結果得られる数値の集合である。データセットは、

10

20

30

40

50

例えば、サンプルの定量化可能な実体を実験的に測定することによって、または代替的に、または研究所、臨床研究組織などのサービスプロバイダーから、または公開もしくは専用データベースから得ることができる。データセットは、データ、およびノードによって表される生物学的実体を含むことができ、該データセットのそれぞれにおけるノードは、同じデータセットにおける、または他のデータセットにおける他のノードと関係していてもよい。さらに、上記ネットワークモデリングエンジン 112 は、例えば、DNA、RNA、タンパク質、または抗体のデータセットにおける遺伝情報から、医療データセットにおける医療情報、患者データセットにおける、また集団全体では、疫学データセットにおける個別の患者に関する情報までを表す計算モデルを生成することができる。上に記載されているさまざまなデータセットに加えて、他の多くのデータセット、または計算モデル

10

【0064】

上記ネットワークモデリングエンジン 112 は、例えば、遺伝子間の調節相互作用、タンパク質間の相互作用、または細胞もしくは組織における複雑な生化学的相互作用を表す 1 つまたは複数のネットワークモデルを生成することができる。上記ネットワークモデリングエンジン 112 によって生成されたネットワークは、静的モデルおよび動的モデルを含むことができる。上記ネットワークモデリングエンジン 112 は、任意の適用可能な数学的スキームを使用して、ハイパーグラフおよび重みづけ二部構成グラフ (weighted bipartite graph) などの、システムを表すことができ、そこでは、ノードの 2 つの種類が反応および化合物を表すために使用される。上記ネットワークモデリングエンジン 112 は、発現量に差のある遺伝子内の機能関係遺伝子の過剰表現に基づく解析、ベイジアンネットワーク解析、グラフィカルガウスモデル技術、または遺伝子関連性ネットワーク技術などの他の推論技術を用いてネットワークモデルを生成して、実験データの組 (例えば、遺伝子発現、代謝産物濃度、細胞応答など) に基づき関連する生物学的ネットワークを識別することもできる。

20

【0065】

上に記載されているように、上記ネットワークモデルは、生物系の機能的特徴の基礎をなす機構および経路に基づく。上記ネットワークモデリングエンジン 112 は、作用物質の長期的な健康上のリスクまたは健康上の利益の研究に関連する生物系の特徴に関する結果を表すモデルを生成するか、または含むことができる。したがって、上記ネットワークモデリングエンジン 112 は、細胞機能、特に、限定はしないが、細胞増殖、細胞性ストレス、細胞再生、アポトーシス、DNA 損傷 / 修復、または炎症応答を含む、生物系内の目的の特徴に関係するか、または寄与する機能の、さまざまな機構についてのネットワークモデルを生成するか、または含むことができる。他の実施形態では、上記ネットワークモデリングエンジン 112 は、急性全身毒性、発がん性、皮膚透過、心血管疾患、肺疾患、生態毒性、目の洗浄 / 腐食、遺伝毒性、免疫毒性、神経毒性、薬物動態、薬物代謝、器官毒性、生殖および発達毒性、皮膚刺激 / 腐食、または皮膚感作性に関連する、計算モデルを含むか、または生成することができる。一般的に、上記ネットワークモデリングエンジン 112 は、核酸 (DNA、RNA、SNP、siRNA、miRNA、RNAi)、タンパク質、ペプチド、抗体、細胞、組織、器官、および任意の他の生物学的実体のステータス、ならびにそれらの各相互作用についての計算モデルを含むか、または生成することができる。一例において、計算ネットワークモデルは、免疫応答または炎症反応の間の免疫系のステータスおよびさまざまな種類の白血球の機能を表すために使用されうる。他の例において、計算ネットワークモデルは、心血管系の性能ならびに内皮細胞の機能および代謝を表すために使用するとも可能である。

30

40

【0066】

本開示のいくつかの実装では、上記ネットワークは、生物学的因果関係知識のデータベ

50

ースから引き出される。このデータベースは、異なる生物学的機構の実験研究を実施して、そのいくつかが因果関係であってもよい機構間の関係（例えば、活性化または阻害関係）を抽出することによって生成され、Cambridge, Massachusetts, USAのSelventa Inc.によって精選された、Genstruct Technology PlatformまたはSelventa Knowledgebaseなどの、市販のデータベースと組み合わせることができる。生物学的因果関係知識のデータベースを使用することで、上記ネットワークモデリングエンジン112は、攪乱102および測定可能要素104 (measurable104)をリンクするネットワークを識別することができる。いくつかの実装では、上記ネットワークモデリングエンジン112は、SRPエンジン110からのシステム応答プロファイルと文献においてすでに生成されているネットワークとを使用して生物学的実体間の因果関係を抽出する。他の処理ステップのうちで、上記データベースをさらに処理して、論理的矛盾を取り除き、生物学的実体の異なる組の間に相通的推論を適用することによって新しい生物学に関する知識を生み出すことができる。

10

【0067】

いくつかの実装では、上記データベースから抽出されたネットワークモデルは、逆因果的推論(RCR)に基づくが、これは因果関係のネットワークを処理して機構仮説を立て、次いで、示差測定結果のデータセットに対してそれらの機構仮説を評価する自動化推論技術である。それぞれの機構仮説は、生物学的実体を、その実体が影響を及ぼしうる測定可能な量にリンクさせる。例えば、測定可能な量として、とりわけ、生物学的実体の濃度、個数、または相対存在量の増減、生物学的実体の活性化もしくは阻害、または生物学的実体の構造、機能、または論理の変化が挙げられ得る。RCRでは、計算のための基盤として生物学的実体間の実験的に観察される因果的相互作用の有向ネットワークを使用する。上記有向ネットワークは、生物学的実体間の相互関係を記録するための構文である、Biological Expression Language (商標) (BEL (商標))で表すことができる。上記RCRの計算では、限定はしないが、経路長(上流ノードと下流ノードとを接続するエッジの最大数)などのネットワークモデル生成、および上流ノードを下流ノードに接続する可能な因果的経路に対するいくつかの制約条件を指定する。RCRの出力は、関連性および精度を評価する統計量によって順位化された、実験測定結果の差異の上流制御機構(upstream controller)を表す機構仮説の組である。したがって、いくつかの実装では、本開示の有用なネットワークモデルは1つまたは複数の機構仮説を含む。上記機構仮説出力をアセンブルして、因果連鎖およびより大きなネットワークを形成し、相互接続されている機構および経路のより高いレベルで上記データセットを解釈することができる。

20

30

【0068】

機構仮説の一種は、潜在的原因を表すノード(上流ノードまたは制御機構)と測定された量を表すノード(下流ノード)との間に存在する因果関係の組を含む。この種類の機構仮説は、上流ノードによって表されている実体の存在量が増えた場合に、因果増大関係によってリンクされる下流ノードが増大すると推論され、因果減少関係によってリンクされる下流ノードが減少すると推論されるなどの、予測を行うために使用することができる。

40

【0069】

機構仮説は、測定されたデータ、例えば、遺伝子発現データの組と、それらの遺伝子の公知の制御機構である生物学的実体との間の関係を表す。それに加えて、これらの関係は、上流実体と下流実体(例えば、下流遺伝子)の差示的発現との間の影響の符号(正または負)を含む。機構仮説の下流実体は、文献で精選されている生物学的因果関係知識のデータベースから引き出されうる。いくつかの実装では、計算可能な因果関係ネットワークモデルの形態の、上流実体を下流実体にリンクする機構仮説の因果関係は、上記NPASコア化法によるネットワーク変化の計算のための基盤である。

【0070】

いくつかの実施形態では、生物学的実体の複雑な因果関係ネットワークモデルは、該モ

50

デルにおける生物系のさまざまな特徴を表す個別の機構仮説を収集し、すべての上記下流実体（例えば、下流遺伝子およびそれらの測定可能な発現レベル）と単一の上流実体または過程との接続を再編成することによって単一の因果関係ネットワークモデルに変換され、これにより、複雑な因果関係ネットワークモデル全体を表すことができ、これは本質的に基礎をなすグラフ構造の平坦化である。したがって、ネットワークモデルで表されているような生物系の特徴および実体の変化は、個別の機構仮説を組み合わせることによって評価することができる。

【0071】

いくつかの実装では、システム100は、細胞が紙巻きタバコの煙、ニコチンを含むエアロゾル、タバコを加熱することによって発生したエアロゾル、またはタバコを燃焼させることによって発生したエアロゾルに曝露されたときの細胞増殖の機構に対するコンピュータ化されたモデルを含むか、または生成することができる。このような一例では、上記システム100は、限定はしないが、がん、肺疾患、および心血管疾患を含む、紙巻きタバコの煙の曝露に関連するさまざまな健康状態を表す1つまたは複数のネットワークモデルを含むか、または生成することもできる。いくつかの態様において、これらのネットワークモデルは、適用される攪乱（例えば、作用物質への曝露）、さまざまな条件の下での応答、目的の測定可能な量、調査されている転帰（例えば、細胞増殖、細胞性ストレス、炎症、DNA修復）、実験データ、臨床データ、疫学データ、および文献のうちの少なくとも1つに基づく。

【0072】

図示されている一例として、上記ネットワークモデリングエンジン112は、細胞性ストレスのネットワークモデルを生成するように構成されうる。上記ネットワークモデリングエンジン112は、文献データベースから公知のストレス応答に関わる関連する機構を記述するネットワークを受け取ることができる。上記ネットワークモデリングエンジン112は、肺および心血管の環境でのストレスに応答して動作することが公知の生物学的機構に基づいて1つまたは複数のネットワークを選択することができる。いくつかの実装では、上記ネットワークモデリングエンジン112は、生物系内の1つまたは複数の機能単位を識別し、より小さなネットワークをそれらの機能性に基づいて組み合わせることによってより大きなネットワークモデルを組み立てる。特に、細胞性ストレスモデルについては、上記ネットワークモデリングエンジン112は、酸化的ストレス、遺伝毒性ストレス、低酸素ストレス、浸透ストレス、生体異物ストレス、および剪断応力への応答に係する機能単位を考慮することができる。したがって、細胞性ストレスモデルに対するネットワーク成分（network component）は、生体異物代謝応答、遺伝毒性ストレス、内皮剪断応力、低酸素応答、浸透ストレス、および酸化的ストレスを含みうる。上記ネットワークモデリングエンジン112は、特定の細胞群において実施されたストレス関連実験からの公に入手可能なトランスクリプトームデータの計算解析からの内容を受け取ることができる。

【0073】

生物学的機構のネットワークモデルを生成するときに、ネットワークモデリングエンジン112は、1つまたは複数のルールを含むことができる。このようなルールは、ネットワーク内容、ノードの種類などを選択するためのルールを含んでよい。上記ネットワークモデリングエンジン112は、*in vitro*および*in vivo*の実験結果の組み合わせを含む、実験データのデータベース106から1つまたは複数のデータセットを選択することができる。上記ネットワークモデリングエンジン112は、実験データを利用して、文献において識別されているノードおよびエッジを検証することができる。細胞性ストレスのモデリングの例において、上記ネットワークモデリングエンジン112は、疾患のない肺または心血管組織において実験が生理学的に関連するストレスをどれほどうまく表しているかに基づき実験についてのデータセットを選択することができる。データセットの選択は、例えば、表現型ストレスのエンドポイントデータの利用可能性、遺伝子発現プロファイリング実験の統計的厳密さ、および通常の疾患のない肺または心血管の生物

10

20

30

40

50

学との実験の環境との関連性に基づくものとする事ができる。

【 0 0 7 4 】

関連するネットワークの集合を識別した後、上記ネットワークモデリングエンジン 1 1 2 はさらに、これらのネットワークを処理し、精密化することができる。例えば、いくつかの実装では、複数の生物学的実体およびそれらの接続は、グループ化され、新しい 1 つまたは複数のノードによって表されうる（例えば、クラスタリングまたは他の技術を使用して）。

【 0 0 7 5 】

上記ネットワークモデリングエンジン 1 1 2 はさらに、識別されたネットワークにおけるノードおよびエッジに関する記述的情報を含むものとしてよい。上に記載されているように、ノードは、その関連する生物学的実体、該関連する生物学的実体が測定可能な量であるか否かの指示、または該生物学的実体の任意の他の記述子によって記述され、その一方、エッジは、例えば、エッジが表す関係の種類（例えば、アップレギュレーションまたはダウンレギュレーション、相関、条件付き依存性、または非依存性などの因果関係）、その関係の強さ、またはその関係における統計的信頼度によって記述されうる。いくつかの実装では、それぞれの処置について、測定可能な実体を表すそれぞれのノードは、上記処置に応答する活性の変化の予測される方向（つまり、増加または減少）に関連付けられている。例えば、気管支上皮細胞が、腫瘍壊死因子（TNF）などの作用物質に曝露される場合、特定の遺伝子の活性が増大しうる。この増大は、文献から公知である（またネットワークモデリングエンジン 1 1 2 によって識別されたネットワークのうちの 1 つで表される）直接的調節関係があるため、またはネットワークモデリングエンジン 1 1 2 によって識別されたネットワークのうちの 1 つまたは複数のエッジを通じて多数の調節関係（例えば、自己分泌シグナリング）を追跡することによって生じうる。いくつかの場合において、上記ネットワークモデリングエンジン 1 1 2 は、上記測定可能な実体のそれぞれについて、特定の攪乱に反応して、変化の予測される方向を識別することができる。上記ネットワークにおける異なる経路が特定の实体についての変化の相反する予測される方向を示す場合、それら 2 つの経路は、変化の正味の方向を決定するためにさらに詳しく調査されうるか、またはその特定の实体の測定結果が破棄されうる。

【 0 0 7 6 】

本明細書において提示されている計算方法およびシステムは、実験データおよび計算ネットワークモデルに基づき N P A スコアを計算する。計算ネットワークモデルは、システム 1 0 0 によって生成されるか、システム 1 0 0 内にインポートされるか、またはシステム 1 0 0 内で（例えば、生物学的知識のデータベースから）識別されうる。ネットワークモデル内の攪乱の下流の効果として識別される実験測定値は、ネットワーク特有の応答スコアの生成において組み合わされる。したがって、ステップ 2 1 6 で、ネットワークスコア化エンジン 1 1 4 は、ネットワークモデリングエンジン 1 1 2 によってステップ 2 1 4 で識別されたネットワークおよび S R P エンジン 1 1 0 によってステップ 2 1 2 で生成された S R P を使用してそれぞれの攪乱について N P A スコアを生成する。N P A スコアは、生物学的実体（識別されたネットワークによって表される）の間の基礎をなす関係の状況において処置（S R P で表される）への生物学的応答を定量化する。ネットワークスコア化エンジン 1 1 4 は、ネットワークモデリングエンジン 1 1 2 内に含まれるか、またはネットワークモデリングエンジン 1 1 2 によって識別されたネットワークのそれぞれについて N P A スコアを生成するためのハードウェア構成要素およびソフトウェア構成要素を備えることができる。

【 0 0 7 7 】

ネットワークスコア化エンジン 1 1 4 は、攪乱に対するネットワークの応答の大きさおよびトポロジー分布を示すスカラー値またはベクトル値のスコアを生成する技術を含む、一群のスコア化技術のうちのいずれかを実装するように構成されうる。一般に、攪乱距離（*p e r t u r b a t i o n m e t r i c*）は、ある刺激または外部事象によりネットワークのモデルにおいて誘発される攪乱を定量化するものである。これらの攪乱距離は、

実験的刺激により生物学的モデル、または他のネットワーク（交通ネットワーク、コンピュータネットワークなど）に誘発される攪乱を定量化するのに特に有用でありうる。攪乱距離は、2つの要素に基づいて生成される。第1の要素は、計算ネットワークモデルであり、対象のシステムの根底にある因果関係ネットワーク（例えば、科学文献で明らかにされている生物学的機構に基づく生物学的ネットワークモデル）に関する任意の知られているデータに基づいて集約することができる。第2の要素は、攪乱が対象のシステムに適用されたときのネットワークモデルの一部または全部の構成要素の挙動を記述する発現データセットである。具体的には、本明細書で使用される場合、発現ノードは通常、発現データが入手可能な計算ネットワークモデルのノードを指す。生物学的解析設定における攪乱解析のいくつかの実施形態では、ネットワークモデルは、生物学的関係の精選された組から構築され、発現データセットは、制御された攪乱が適用され監視される実験によって生成される。ネットワークのトポロジーを明確に使用して、ネットワークの最も攪乱されそうな領域または特定の領域を識別する攪乱解析方法が本明細書に記載される。

10

【0078】

一例では、攪乱距離は、対応するノードにおける2つのデータセット（つまり、処置データセットとコントロールデータセット）の間の差（または倍率変化値）を表す。この攪乱距離は、攪乱指標とすることができ、対応するノードの活性が攪乱から影響を受ける程度を表すことができる。具体的には、図6に関連してより詳細に記載されているように、攪乱指標は、所与のノードから下流のノードの測定された活性の一次結合として計算することができる。

20

【0079】

ネットワークモデルは、エッジを越えて相互接続されるノードを含み、ネットワークモデル内のエッジは遷移確率と関連付けることができる。遷移確率は、ネットワーク内の1つのノードから別のノードへの遷移の尤度を表すことができる。一例として、遷移確率は、対応するノードにおける2つのデータセット（つまり、処置データセットとコントロールデータセット）の間の差を表す攪乱距離に少なくとも一部は基づいて計算される。一例として、図7に関連してより詳細に記載されているように、遷移確率は、ノードの攪乱指標の一次関数として計算することができる。さらに、ネットワーク内のエッジの遷移確率を使用してノード距離（node metric）を決定することができる。対応するノードのノード距離は、ノードの相対的影響を表すことができる。図5に関連してより詳細に記載されているように、ネットワーク内のエッジの遷移確率を計算することに加えて、ネットワーク内のノードの平衡確率もまた計算することができる。対応するノードの平衡確率は、その対応するノードを定常状態においてランダムウォークが訪問する尤度である。

30

【0080】

具体的には、ネットワーク内のノードの中心度値を計算して、ネットワーク内のノードの相対的重要度を表すことができる。ネットワーク内のあるノードの相対的重要度は、ネットワーク内のそのノードと他のノード間の関係を表せるものであり、そのネットワークにおける遷移確率、平衡確率、または遷移確率と平衡確率の両方に依存しうる。一例として、ネットワークを通る横断がランダムウォークモデルによって表される場合、ランダムウォークで頻繁に訪問されるノードは、あまり頻繁に訪問されない他のノードよりも相対的に重要でありうる。すなわち、より頻繁に訪問されるノードは、より大きい中心度値を有し、あるノードについての中心度値の計算は、他のノードへの連続する訪問の間の、対応するノードへのランダムウォークの予想される訪問の回数に基づくことができる。具体的には、図8に関連してより詳細に記載されているように、中心度値は、ネットワーク内のノードすべてにわたって予想される訪問の回数の一次結合として計算することができる。一例では、中心度値の計算は、遷移確率が下流ノードの測定された活性レベルをベースとする「強化」ランダムウォークモデルに基づいている。

40

【0081】

ネットワーク内のノードの中心度値を使用して、ネットワークのトポロジー全体を調べ

50

ることができる。一例では、ネットワーク内の1つのノードにおける攪乱が別のノードの中心度値に影響を及ぼすことがある場合に、感度解析を行うことができる。このようにして、ネットワークのトポロジーは、ネットワークの1つの位置においての、別の位置の変化の影響を理解するために使用される。別の例では、ネットワーク内のノードの中心度値を使用して、ネットワーク全体にわたる攪乱のトポロジーを視覚化することができる。具体的には、中心度値をスペクトル変換を用いて射影し、射影のサブセットを表示することによりノイズが軽減され得、その結果、ネットワーク内の重要な経路を容易に視覚化できるようになる。別の例では、ネットワーク内のノードの中心度値を集約して、攪乱に対するネットワークモデルの全体応答を表すスカラー値を定義することができる。一般に、ネットワーク内のノードの中心度値を使用して、ネットワークに対する種々の攪乱のあらゆるトポロジー効果を調べることも視覚化することもできる。

10

【0082】

図5～図8は、ネットワーク内のノードにおける攪乱と、ネットワーク内の異なるノード間の遷移と、ネットワーク内のノードの中心度値とに関連する値を生成するための例示的な方法の流れ図である。加えて、図4Bおよび図11は、上流ノード、下流ノードおよびエッジを含む例示的なネットワークの図であり、図5～図8の流れ図と関連して描かれている。具体的には、図5の流れ図は、ネットワーク内のあるノードの相対的重要度の尺度に相当するノードの中心度値を計算するための包括的な方法である。図6～図8に示されているプロセスは、図5の流れ図の種々のステップで使用することができる。具体的には、図6の流れ図は、ある選択されたノードの攪乱指標を計算するための一方法である。攪乱指標とは、選択されたノードから下流のノードの活性レベルと関連付けられた値である。加えて、攪乱指標は、ネットワーク内の異なるノードを接続するエッジが改変される「強化」ランダムウォークモデルの判定に使用することもできる。強化ランダムウォークモデルについては、図7に関連してより詳細に記載される。最後に、図8の流れ図は、強化ランダムウォークモデルに基づいて中心度値を計算するための方法である。

20

【0083】

図5は生物学的ネットワーク内のノードの中心度値を生成する例示的なプロセス500の流れ図である。上に記載されているように中心度値は、ネットワーク内のノードの相対的重要度を表す。ステップ502で、対象のシステムの因果関係ネットワークモデルが識別される。図1および図2に関連して上に記載されたように、ネットワークモデリングエンジン112は、1つまたは複数のデータセットを併合しやすくする、あるいは1つまたは複数のネットワークを併合しやすくすることによって、モデルの一部を受け取ること、および/または生成することができる。有向ネットワークGは、因果関係ネットワークモデルの根底にあるネットワークである。ネットワーク内のn個のノード(例えば、生物学的実体、交通位置、ソーシャルネットワーク内の個体を表す)は、 $(V_i)_{i=1, \dots, m}$ で表示される。有向ネットワーク $G = (V, E)$ は、次式により定義される隣接行列Aで表すことができる。

30

【化1】

$$A_{ij} = \begin{cases} 1 & i \rightarrow j \text{ の場合} \\ 0 & \text{そうでない場合} \end{cases} \quad (1)$$

40

具体的には、有向エッジが第1のノードiから第2のノードjまで存在する場合、隣接行列Aの要素は1になる。そうでなければ、隣接行列Aの要素は0になる。Iがノードの組を表すものとするが、このノードの組には、実験データをマッピングすることができる他のノードが(上流または下流に)ある。実験データをマッピングすることができるノードは、発現ノードとすることができる。具体的には、ノードの組Iは、ネットワーク内のm個すべてのノードの任意のサブセットを含むことができる。図11はこのようなシナリオを示し、ネットワーク内の4つのノード1102a～1102d(全体ではノード1102)が提示されている。加えて、遺伝子チップ1106が複数のプローブセット1104を含み、各プローブセット1104の斜線パターンおよび位置は、ある特定の遺伝子の発

50

現レベルを表している。各ノード1102は、下流遺伝子1108a~1108c(全体で下流遺伝子1108)の組を有し、矢印は、下流遺伝子1108と複数のプローブセット1104のサブセットとの間の関連を示す。図を分かりやすくするために、図11では、下流遺伝子1108の1つのサブセットおよびプローブセット1104だけが標示されている。具体的には、図11に例示されたシナリオは、因果モデルと実験データの間のリンクを示している。

【0084】

ステップ504で、少なくとも1つの下流の測定可能ノードまたは発現ノードを用いて、攪乱指標(PI)がIの中のノードごとに生成される。具体的には、あるノードのPIは、そのノードからの下流活性量を表す。具体的には、図6に関連して以下でより詳細に記載されるように、上流ノードと下流ノードの間に因果関係が存在する場合、下流ノードは、上流ノードの活性に関する支持証拠を与えることができる。図11の例示的ネットワーク1100では、上流ノード1102は下流ノード1108との因果関係を有する。すなわち、上流ノード1102aのPIは、下流ノード1108における活性レベルに依存する。

10

【0085】

一例では、PI値は、ノード1102の活性(例えば、遺伝子相互作用ネットワーク、またはタンパク質-タンパク質相互作用ネットワークによって表される生物系内の転写の数)が、ネットワーク1100内の別の位置で適用された攪乱から影響を受ける程度を表す。ノードのPIは、根底にある機構が活性化された(阻害または増強された)という証拠についての情報をもたらす。攪乱が実験設定に適用された場合、ノードの活性は、対照条件におけるノードの活性と処置条件におけるノードの活性との間の相対的測定値とすることができる。

20

【0086】

図6は、ある選択されたノードのPIを決定する例示的プロセス600の流れ図である。プロセス600は、例えば、ネットワークスコア化エンジン114、またはシステム100の構成要素のうちの適切に構成された他の任意の構成要素によって実装することができる。図6に描かれているように、選択されたノードのPIを決定することには、選択されたノードから下流のノードの活性尺度の一次結合を計算することが含まれる。ステップ602で、ネットワークスコア化エンジン114は、ノードの組Iの中でノードiを選択する。一例では、ネットワークスコア化エンジン114は、ネットワーク1100内のノード1102aを選択する。

30

【0087】

ステップ604で、ネットワークスコア化エンジン114は、ステップ602で選択されたノード1102aから下流のノードを識別する。下流ノードは、選択されたノードiの下流の発現ノードでありうると共に、遺伝子発現(または測定可能ノード1104、この場合、測定可能ノード1104のパターンは、測定された活性レベルの値に対応しうる)を表すことができる。下流ノードは、上記の式1で定義された隣接行列Aによって定義される因果関係ネットワークモデルに基づいて識別することができる。具体的には、識別された下流ノードは、識別された下流ノードが、選択されたノード1102aの直接隣接ノードになるように、単一の有向エッジ(またはリンク)と共に、選択されたノードiからすべて分離することができる。加えて、識別された下流ノードは、対応する測定可能ノード1104を有する選択されたノード1102aの直接の下流隣接ノードに対応しうる。

40

【0088】

ステップ606で、ネットワークスコア化エンジン114は、識別された下流ノード1108(ステップ604で識別)の、別々の処置条件に対する活性変化を決定する。具体的には、この活性変化は、コントロールデータと処置データの間で、または別々の処置条件を表す2組のデータ間で、あるノード測定値が初期値から最終値までどれだけ変化するかを記述する数の実験結果でありうる。具体的には、識別された下流ノードkについて、

50

その活性変化は、ノードkの倍率変化 v_k によって表すことができる。具体的には、 v_k の正の値は、処置データの結果としてノードkにおける活性増加を表すことができ、 v_k の負の値は活性減少を表すことができ、あるいはその逆とすることができる。いくつかの実施形態では、活性変化は、2つの条件間での生物学的実体の活性の倍率変化の対数とすることができる。一般に、倍率変化 v_k は、ノードkの活性化の他の任意の指標（絶対的または相対的）を表しうる。

【0089】

ステップ608で、ネットワークスコア化エンジン114は、ステップ604で識別された下流ノード1108の局所的偽非発見率 (local false non-discovery rate) ($fndr$) を決定する。具体的には、局所的偽非発見率 $fndr$ (つまり、場合によっては、観察されたp値を条件として、倍率変化値 v_k がゼロ倍率変化の基礎となる帰無仮説からの逸脱を表す確率) であり、これはStrimmerら、「A general modular framework for gene set enrichment analysis」、BMC Bioinformatics 10:47、2009年およびStrimmer、「A unified approach to false discovery rate estimation」、BMC Bioinformatics 9:303、2008年に記載されているとおりであり、それぞれ参照によりその全体が本明細書に組み込まれている。言い換えると、 $fndr$ は、倍率変化値 v_k が0と有意に異なって、別々の処置条件を表す2つのデータセット間に有意差があることを意味する確率を表すのに使用することができる。高い $fndr$ は、別々の処置条件の結果としてデータにおける有意差が生じたことを意味する。局所的 $fndr$ は、偽発見率 fdr (つまり、倍率変化値 v_k がゼロ倍率変化の基礎となる帰無仮説からの逸脱を表さない確率) に基づくことができる。具体的には、局所的 $fndr$ は、下流ノードkについて、 $fndr_k = 1 - fdr_k$ によって定義することができる。一例では、偽発見率 fdr_k は、調整されたp値 (つまり、ゼロ倍率変化の帰無仮説が真であると想定して、実際に観察された倍率変化 v_k ほどに少なくとも極端な倍率変化が得られる確率) に少なくとも依存する。

【0090】

ステップ610で、ネットワークスコア化エンジン114は、選択されたノードi (つまり、ノード1102a) の攪乱指標PIを計算する。具体的には、 PI_i は、識別された下流ノード (つまり、ノード1108) の活性変化および偽非発見率に基づいて計算することができる。一例では、 PI_i は、活性変化と偽非発見率の集合尺度 (aggregate measure) になりうる。一例として、ネットワークスコア化エンジン114は、 PI_i を下流ノードの $fndr$ と v_k の絶対値に基づく発現の一次結合として次式により計算することができる。

【化2】

$$PI_i = \frac{1}{\| \text{下流ノード } v_k \|} \| fndr \cdot \beta \|_{L^1(\text{下流ノード } v_k)} \quad (2)$$

具体的には、下流ノード1108は、選択されたノード1102aの子ノードであり、ある特定の遺伝子の発現の特別な形のものである。これらの子ノードは、実験データに直接リンクされているものである。ノード1108などの下流ノードでは、 $fndr$ と倍率変化 v_k の間の積は、別々の処置条件から生じたデータセットにおける差の変倍バージョンを表す。式2で、ネットワークスコア化エンジン114は、ノードiの下流ノード全体にわたり、これらの変倍された倍率変化値の絶対値の平均として PI_i の値を計算する。変倍された倍率変化値は、下流ノードの活性尺度を表す。一般に、 PI_i は、これらの変倍された倍率変化値の一次結合として、下流ノード全体にわたって計算することができる。したがって、大きく有意な倍率変化 v_k を有する下流ノードの場合、下流ノードは、上流ノードiの PI_i に対し大きい値をもたらす。式2は、適用された攪乱からノードの活性が影響を受ける程度を表すノードのPIを計算する一方法である。具体的には、PIは、倍率

変化値に依存する幾何学的攪乱指標 (GPI) (Geometric Perturbation Index) スコアとすることができ、これは、Martinらの BMC systems biology 2012, 6:54、および係属中の特許出願 PCT/EP2012/061035 に記載されているとおりであり、これらは両方とも参照によりその全体が本明細書に組み込まれている。しかし、一般に、任意の適切な尺度をノードのPIとして使用することができる。

【0091】

次に図5に戻ると、ステップ506で、ネットワークスコア化エンジン114は、ネットワークGについて強化ランダムウォークを定義する。強化ランダムウォークでは、ある特定の因果関係と関連付けられた遷移確率は、下流PIに依存する(もしあれば)。例示的な一例として、図4Bは、ノード412a~412d(全体ではノード412)およびエッジ410a~410b(全体ではエッジ410)を含むネットワーク400bの図である。図を分かりやすくするために、ネットワーク400bでは、ノードとエッジの1つのサブセットだけが標示されている。エッジ410は、エッジによって接続された2つのノード間の遷移が、矢印で示された1つの方向に起こることを示すように方向付けられる。一例として、エッジ410aに対して、ノード412aは上流ノードとみなすことができ、ノード412bは下流ノードとみなすことができる。ノード412aと412bの間の因果関係を強化するために、ノード412aからノード412bへの遷移の確率は、412bのPI値に依存している。さらには、ノード412bのPI値は、ノード412dなどの、ノード412bからさらに下流のノードの測定された活性レベルに依存している。こうして強化ランダムウォークは、下流ノードのPIに基づいて因果ステートメントを強化する。ランダムウォーク中により横断されやすいノードはネットワークの中心のノードになるので、強化ランダムウォークの解析により、モデルの各ノードの重要度についての情報が得られる(つまり、因果律(causality)の流れはノードの重要度に関係する)。

【0092】

ステップ506で定義された強化ランダムウォークについての記載が、いくつかの事前の注釈および説明の後で行われる。ネットワークG上のランダムウォークは、離散時間マルコフ連鎖によって表すことができ、その状態空間はV(ネットワークのノードセットまたは頂点セット)であり、また遷移確率 p_{ij} が $A_{ij}=0$ であれば $p_{ij}=0$ に制約される。遷移確率 p_{ij} は、ノードiからノードjに移動するランダムウォークの確率を表す。このマルコフ連鎖は、 $M_{ij}=p_{ij}$ で定義される遷移行列M(順方向伝搬演算子とも呼ばれる)によって表すことができる。この行列は確率的であり、頂点セットについての初期確率分布と共に、ネットワークについての離散時間マルコフ連鎖 $(X_n)_{n=0}$ を完全に定義する。ネットワーク内のエッジによって表されるネットワークトポロジーおよび因果律を考慮して、伝搬演算子Mは、ノード間の因果関係を通して進展するランダムウォークを規定する。

【0093】

あるマルコフ連鎖が非周期的であり既約な場合、このマルコフ連鎖は、次式により定義される平衡測度(つまり、平衡確率)を有する。

$$M = \quad (3)$$

具体的には、平衡測度は、長さmのベクトル(mはネットワーク内のノードの数)である。平衡測度中の各要素は、ネットワーク内のノードに対応し、その対応するノードを定常状態において訪問するランダムウォークの全確率になる。定常状態(または平衡)に達した後、どのノードを訪問するランダムウォークの確率もやがて固定される。

【0094】

平衡測度は、初期分布を表す任意の測度 μ について、 n のとき μM^n がに収束するという観察結果を用いて、反復法によって計算することができる。ここでnは時間を表す整数である。具体的には、 M^n は、すべてのノードiについて

10

20

30

40

【化 3】

$$M_{ij}^{\infty} = \pi_j$$

を満たす階数 1 の行列 M に指数関数的に速く収束する。エルゴードの定理によれば、

【化 4】

$$N_n^{(i)}$$

が時刻 n の前のノード i への訪問回数を表す場合、任意の初期分布について、 n のとき確率 1 で

【化 5】

$$\frac{N_n^{(i)}}{n} \rightarrow \pi_i$$

である。図 8 に関連してより詳細に記載されるように、平衡測度 を使用して、ネットワーク内のあるノードの相対的重要度を計算し、そうして、そのノードの中心度値を計算することができる。

【0095】

ネットワークスコア化エンジン 114 はまた、ノード i をランダムウォークで訪問する最初の時刻に対応する第 1 の到達時間を定義することもできる。具体的には、ノード i への第 1 の正の到達時間は

【化 6】

$$T_i^+$$

で示され、次式により計算することができる、

【化 7 - 1】

$$T_i^+ = \min\{n \geq 1 | X^n = i\}, \quad (4)$$

一方、ノード i への第 1 の到達時間は T_i で示され、次式により計算することができる。

【化 7 - 2】

$$T_i = \min\{n \geq 0 | X^n = i\}. \quad (5)$$

図 8 に関連してより詳細に記載されるように、第 1 の正の到達時間

【化 8】

$$T_i^+$$

および第 1 の到達時間 T_i を使用して、ネットワーク内のノードの中心度値を計算することができる。

【0096】

有限エルゴード的マルコフ連鎖の基本行列またはグリーンの測度は、次式により定義することができる、

【化 9 - 1】

$$G = \sum_{n \geq 0} (M^n - M^{\infty}), \quad (6)$$

または同様に、次式により定義することができる。

【化 9 - 2】

$$G_{ij} = \sum_{n \geq 0} (p_{ij}^n - \pi_j), \quad (7)$$

ここで

【化 10】

$$p_{ij}^n$$

10

20

30

40

50

は、ノード i から出発したランダムウォークが n ステップの後にノード j にある確率である。一般に、時刻 0 と t の間にランダムウォークがノード j で費やす平均時間量は、出発ノード i に関係なく、おおまかに $(t + 1) \pi_j$ と推定することができる。しかし、出発ノード i が分かっている場合は、グリーンの測度 G_{ij} は、おおまかな推定値と組み合わせられるべき補正項を表す。具体的には、 $G_{ij} = \lim_{t \rightarrow \infty} (T_{ij}(t) - (t + 1) \pi_j)$ であり、ここで $T_{ij}(t)$ は、ノード i を出発するランダムウォークが時刻 0 と t の間にノード j を訪問する平均回数 (average number of times) に相当する。図 8 に関連してより詳細に記載されるように、マルコフ連鎖の基本行列を使用して、ネットワーク内のノードの中心度値を計算することができる。

【0097】

10

【化11】

$$G_i \doteq \sum_{n \geq 0} (\delta_i - \pi) M^n$$

が演算子

【化12】

$$\mu \mapsto \mu M + (\delta_i - \pi)$$

の不動点であるので、この不動点は、ノード i のソース 1 および一様シンク (uniform sink) - を連続して与えるソース項 π_i を用いて、ランダムウォークの平衡測度として表すことができる。結果として、量 G_i は、ノード i のソースと共にページランクによって表すことができる。

20

【0098】

以下のリストは、 G および G の例示的特性を列挙するものである。これらおよび他の特性は、Aldous および Fill の Reversible Markov Chains and Random Walks on Graphs でさらに詳細に記載されており、これは <http://www.stat.berkeley.edu/~aldous/RWG/book.html> で入手可能であり、参照によりその全体が本明細書に組み込まれている。表記

【化13】

$$E_\mu(\cdot)$$

30

は、初期分布 μ の予想値を示す。表記

【化14】

$$E_i(\cdot)$$

は、初期分布 π_i の予想値を示す。

【化15】

i) 全てのノード j について $\sum_j G_{ij} = 0$ であり、通常 G は自己共役ではない；

ii) $E_i(T_i^+) = \frac{1}{\pi_i}$;

iii) $E_i(\text{時間 } T_i^+ \text{ 前の } j \text{ への訪問数}) = \frac{\pi_j}{\pi_i}$;

40

iv) $\pi_i E_\pi(T_i) = G_{ii}$;

v) $\pi_j E_\pi(T_j) = G_{jj} - G_{ij}$;

vi) $E_\pi(\text{時間 } T_i \text{ 前の } j \text{ への訪問数}) = \frac{\pi_j}{\pi_i} G_{ii} - G_{ij}$.

【0099】

ステップ 506 で定義された強化ランダムウォークは、より大きい PI を有するノードへ向かって遷移が有利になっているランダムウォークである。強化されないランダムウォークの一例として、ネットワーク内のすべてのエッジが同じ遷移確率を有し得る。しかし

50

、強化ランダムウォークでは、遷移選好はPI、またはPIの一次関数に比例しうる。具体的には、特定の因果関係（つまり、ネットワーク400b内のエッジ410a）と関連付けられた遷移確率は、下流ノード（つまりノード412b）のPIに依存する。したがって、強化ランダムウォークは、下流ノードのPIに基づいて因果ステートメントを強化する。したがって、強化ランダムウォークの解析により、ランダムウォーク中により横断されやすいノード（つまり、高確率の入ってくるエッジを伴うノード）に関する、すなわちネットワークの中心の重要なノードに関する情報が得られる。

【0100】

いくつかの実施形態では、ネットワークスコア化エンジン114は、図7の方法700を使用して、ステップ506の強化ランダムウォークの伝搬演算子 $M^{-1} \cdot V$ を計算することができる。具体的には、伝搬演算子Mは、要素がノード間の遷移確率に対応する行列である。図7に描かれているように、行列Mの要素は、ノードPI値の一次関数である。具体的には、dがノードiから出て行くエッジの数（つまりノードiの外れ度）である場合、伝搬演算子Mは次式により定義することができる。

10

【化16】

$$M_{ij} \propto \begin{cases} \frac{1}{d} (1 + 100 \cdot PI_j) & i \rightarrow j \text{ および } j \in I \text{ の場合} \\ \frac{1}{d} & i \rightarrow j \text{ および } j \notin I \text{ の場合} \\ 0 & \text{そうではない場合} \end{cases} \quad (8)$$

20

次に図7を参照すると、プロセス700は、式8により伝搬演算子Mの要素 M_{ij} を決定するために、ネットワークスコア化エンジン114によって実装することができる。ステップ702で、ネットワークスコア化エンジン114は、2つのノードi（つまり、ノード412a）とj（つまり、ノード412b）の間の遷移を選択する。具体的には、ネットワーク内の任意の2つのノードを選択することができ、かつ1つの方向を選択することができる。判断ブロック704で、ネットワークスコア化エンジン114は、有向エッジ $i \rightarrow j$ （つまり、エッジ410a）が存在するかどうかを判定する。有向エッジが存在しない場合、ノードiからノードjへの遷移の確率が0であるので、ネットワークスコア化エンジン114は、ステップ706で要素 M_{ij} に0の値を割り当てる。有向エッジが存在する場合、ネットワークスコア化エンジン114は判断ブロック708へ移行して、ノードiがノードの組Iの中にあるかどうかを判定する。一例では、ネットワークスコア化エンジン114は、判断ブロック708で、ネットワークモデルを調べてノードiがいずれかの発現ノードに、または実験データをマッピングできる他のいずれかのノードに接続されているかどうか（つまり、上流または下流で）を判定する。具体的には、ノードの組Iは、実験データへの直接リンクを有するノード1102の組である。具体的には、ノードiがノードの組Iの中になく、ネットワークスコア化エンジン114は、ステップ710で、

30

【化17】

$$\frac{1}{n}$$

40

に比例する値を要素 M_{ij} に割り当てる

【化18】

(すなわち、 $M_{ij} \propto 1/n$)

そうでなければ、ネットワークスコア化エンジン114は、ステップ712で、

【化19】

$$\frac{1}{n} (1 + 100 \cdot PI_j)$$

に比例する値を要素 M_{ij} に割り当てる

【化 20】

(すなわち、 $M_{ij} \propto (1 + 100 \cdot PI_j) / n$)

具体的には、要素 M_{ij} の値は、 j すべてにわたる要素 M_{ij} の合計が 1 に等しくなるように正規化することができる。

【0101】

図 7 に示されたプロセス 700 は、PI 値に基づいて遷移を優先的に重みづけすることによって、ネットワーク内の異なるノード間の遷移の確率を改変する実装の一例である。しかし、一般に、遷移確率を改変するには任意の適切な方法を使用することができる。

【0102】

加えて、式 8 の遷移確率によって定義されたマルコフ連鎖は、必ずしも既約ではない。例えば、吸収ノードが存在しうる（細胞活性を表す生物学的ネットワーク内のアポトーシスなど）。一例として、図 12 のネットワーク内のノード N23、N51、N77、N95、N100、および N104 は、入ってくるエッジだけを有し、出て行くエッジを有さない吸収ノードの例である。いくつかの実施形態では、この問題は、ランダムウォークが 1 つまたは複数の指定ノード（例えば、上流ノードがないノード）へ逃れることができるように、追加の遷移確率を含むことによって対処される。いくつかの実施形態では、この問題は、ランダムウォークが一部または全部のノードでランダムジャンプを行うことができるように、追加の遷移確率を含むことによって対処される。

【0103】

次に、図 5 を参照すると、ステップ 508 で、ネットワーク内の個別ノードに対し中心度値が生成される。一般に、あるノードの中心度値により、ネットワーク内のそのノードの相対的重要度を定量化する。例えば、あるノードの中心度値は、ネットワーク内の他のノードに関して定義することができる。具体的には、選択されたノードの中心度値は、強化ランダムウォークが別のノードに初めて訪問する前に、選択されたノードを訪問する予想数に基づいて計算することができる。中心度値の一例は、White および Smyth の Algorithms for estimating relative importance in networks、International Conference on Knowledge Discovery and Data Mining、Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining、2003 年、266 ~ 275 頁に記載されており、その全体が参照により本明細書に組み込まれている。

【0104】

次に図 8 を参照すると、プロセス 800 は、ネットワーク内のノードの中心度値を生成するために、ネットワークスコア化エンジン 114 によって実装することができる。上に記載されたように、ノードの中心度値は、ネットワーク内のあるノードの相対的重要度を表し、またネットワーク内のそのノードと他のノードとの間の関係を表すことができる。加えて、中心度値は、強化ランダムウォークモデルに依存することができる（図 7 に関して伝搬演算子 M について定義されたとおり）。一例では、対応するノードの中心度値は、他のノードへの連続する訪問の間の、対応するノードへのランダムウォークの予想される訪問数に基づいて計算される。こうして、中心度値は、そのノードをランダムウォークが訪問する予想回数 (expected number of times) を表し、したがって、ネットワーク内のそのノードの相対的重要度を示す。

【0105】

具体的には、ステップ 802 で、ネットワークスコア化エンジン 114 は、式 6 および式 7 により基本行列 G を計算する。ステップ 804 で、ネットワークスコア化エンジン 114 は、ノード i への最初の訪問の前のノード j への訪問の予想数を決定する。いくつかの実施形態では、上記の特性リストからの性質 (v_i) が、ステップ 804 で適用される。ステップ 806 で、ネットワークスコア化エンジン 114 は、すべてのノード i にわた

10

20

30

40

50

って訪問の予想数を合計し、ステップ 808 で、ノード j の中心度値が、ステップ 806 で計算された合計に設定される。具体的には、ノード j のマルコフ中心度は次式により計算される。

【化 2 1】

$$C(j) = \sum_{i=1, \dots, m} \left(\frac{\pi_j}{\pi_i} G_{ii} - G_{ij} \right) \tag{9}$$

$$= \sum_{i=1, \dots, m} \mathbb{E}_{\pi}(\text{時間 } T_i \text{ の前の } j \text{ への訪問数}) \tag{10}$$

したがって、ノード j の中心度値は、別のノードを訪問する前にランダムウォークがノード j を訪問することが予想される回数に基づいている。極端な場合では、ランダムウォークが他のノードを初めて訪問する前に 1 つのノード j 1 が何回も訪問されるならば、ノード j 1 は相対的に重要であり、その結果、大きい中心度値 C (j 1) 得られることになる。一方で、ランダムウォークが他のノードを初めて訪問する前にノード j 2 が訪問されないならば、ノード j 2 は相対的に重要でなく、その結果、より小さい中心度値 C (j 2) 得られることになる。

10

【 0 1 0 6 】

いくつかの実施形態では、個々のノード j の中心度値を計算するには、強化ランダムウォーク (ステップ 506 で定義) のマルコフ中心度は、いかなるデータによっても強化されない (つまり、すべてのノード i に対して $P I_i = 0$ である) ランダムウォークについて計算された中心度と組み合わせるとよい。強化されないランダムウォークは、単純ランダムウォーク (SRW) と呼ばれることがあり、強化ランダムウォークと SRW の間の比較により、強化ランダムウォークに PI を含むことの影響を識別することができる。SRW のマルコフ中心度を $C^{SRW}(j)$ で示す。いくつかの実施形態では、中心度値は次式により生成される。

20

【化 2 2】

$$R(j) = \log_{10} \left(\frac{C(j)}{C^{SRW}(j)} \right) \tag{11}$$

強化マルコフ連鎖中心度および SRW の中心度を含む中心度値を使用することによって、対象のシステムの観察された挙動がネットワークモデル内の経路を強化することができる。強化ランダムウォーク内の PI 値のすべてがゼロである場合、すべての j について R (j) がゼロになる。

30

【 0 1 0 7 】

式 9 ~ 式 11 は、あるノードの中心度値を計算するための種々の技法の説明的な例であり、異なる技法により異なる利点をもたらされ得る。例えば、式 11 は、SRW に対して正規化された値としての強化ランダムウォークの中心度値を表し、このようにして不変測度になる。式 10 で記述された予想訪問数アプローチは、PI による強化に対する感度が不変アプローチよりも高い可能性がある。最後に、式 9 で記述されたグリーン測度もまた、中心度値を得るのに使用することができるが、予想訪問数アプローチのような即時の確率解釈は与えない。

【 0 1 0 8 】

40

一般に、本明細書に記載されている技法は、実験データまたは観察データが利用可能なシステムを表すためにネットワークモデルが使用される、あらゆる状況に適用することができる。例えば、エッジが道路容量 (road capacity) によって重みづけされ、各ノードが道路交差点であり、また発現ノードが事故データまたは交通渋滞データを利用可能な道路交差点でよいネットワークによって、交通ネットワークを表すことができる。事故データまたは交通渋滞データを使用してランダムウォークモデルにバイアスをかけ、交通量の変化に応じて道路交差点における挙動を予想することができる。別の例では、エッジがウェブページ間のリンクであり、各ノードがウェブページであり、また発現ノードが訪問者データの利用が可能なページでよいネットワークによって、ウェブネットワークを表すことができる。訪問者データを使用してランダムウォークモデルにバイアスを

50

かけ、ウェブサーフィン習慣の変化に応じてウェブページへの訪問を予測することができる。

【 0 1 0 9 】

図 5 および図 8 で計算されたネットワーク内のノードの中心度値を使用して、ネットワークの全体トポロジを調べることができる。ネットワーク内の中心度値を使用してネットワークのトポロジを調べるための少なくとも 3 つの例示的方法が本明細書に記載される。1 つの例では、ネットワークスコア化エンジン 1 1 4 は、ネットワーク内の 1 つのノードにおける攪乱が別のノードの中心度値に及ぼす影響を調べるための感度解析を行うことができる。このようにして、ネットワークのトポロジが、ネットワークの 1 つの位置における変化の別の位置における影響を理解するために使用される。第 2 の例では、ネットワーク内のノードの中心度値を使用して、攪乱のトポロジをネットワーク全体にわたって視覚化することができる。具体的には、これらの視覚化方法によりノイズが軽減され得、その結果、ネットワーク内の重要な経路を容易に視覚化できるようになる。第 3 の例では、ネットワーク内のノードの中心度値を集約して、攪乱に対するネットワークモデルの全体応答を表すスカラー値を定義することができる。これら 3 つの例が、以下でより詳細に記載される。しかし、一般に、ネットワーク内のノードの中心度値を使用して、ネットワークに対する種々の攪乱のあらゆるトポロジ効果を調べる、または視覚化することができる。

10

【 0 1 1 0 】

いくつかの実施形態では、あるノードの攪乱指標の変化と、別の（または同じ）ノードの中心度値との間の関係を理解するには、ネットワークスコア化エンジン 1 1 4 により感度解析を行うことが所望され得る。ネットワークのより深い解析は、ネットワークノードの中心度値に対する実験的証拠の影響を理解することによって（例えば、PI 値を介して）行うことができる。いくつかの実施形態では、感度解析は以下の数式の値または近似値を決定することを含む。

20

【 化 2 3 】

$$\frac{\partial R(j)}{\partial PI_k} \quad (12)$$

式 1 2 の数式は、次のように書き表すことができる。

30

【 化 2 4 】

$$\frac{\partial C(j)/\partial PI_k}{C(j)} - \frac{\partial C^{SRW}(j)/\partial PI_k}{C^{SRW}(j)} \quad (13)$$

基本行列 G は、次のように表すことができる。

$$G = (I - (M - M))^{-1} - M \quad (1 4)$$

【化 2 5】

加えて、 $\mathbb{E}_{\pi}(\text{時間 } T_i \text{ の前の } j \text{ への訪問数})$ は、次のように表現することができる。

$$\text{diag}(G) \left(\frac{1}{\pi} \right) \pi^T - G. \tag{15}$$

したがって、

$$\frac{\partial G}{\partial P_{I_k}} = -(G + M^\infty) \left(\frac{\partial(I-M)}{\partial P_{I_k}} + 1 \cdot \frac{\partial \pi^T}{\partial P_{I_k}} \right) (G + M^\infty) - \frac{\partial M^\infty}{\partial P_{I_k}} \tag{16}$$

$$= (G + M^\infty) \left(\frac{\partial(M-I-1\pi^T)}{\partial P_{I_k}} \right) (G + M^\infty) - \frac{\partial 1\pi^T}{\partial P_{I_k}} \tag{17}$$

$$= (G + M^\infty) \left(\frac{\partial M}{\partial P_{I_k}} - 1 \cdot \frac{\partial \pi^T}{\partial P_{I_k}} \right) (G + M^\infty) - 1 \cdot \frac{\partial \pi^T}{\partial P_{I_k}} \tag{18}$$

式18の結果を式10の数式と共に使用すると、以下が得られる。

$$\frac{\partial C(j)}{\partial P_{I_k}} = \frac{\partial}{\partial P_{I_k}} \left(\text{diag}(G) \left(\frac{1}{\pi} \right) \pi^T - G \right) \tag{19}$$

$$= \frac{\partial \text{diag}(G)}{\partial P_{I_k}} \left(\frac{1}{\pi} \right) \pi^T + \text{diag}(G) \frac{\partial \left(\frac{1}{\pi} \right)}{\partial P_{I_k}} \pi^T + \text{diag}(G) \left(\frac{1}{\pi} \right) \frac{\partial \pi^T}{\partial P_{I_k}} - \frac{\partial G}{\partial P_{I_k}} \tag{20}$$

ここで、

$$\frac{\partial \left(\frac{1}{\pi} \right)}{\partial P_{I_k}} = - \left(\frac{1}{\pi^2} \right) \cdot \frac{\partial \pi}{\partial P_{I_k}} \tag{21}$$

および

$$\frac{\partial \text{diag}(G)}{\partial P_{I_k}} = \text{diag} \left(\frac{\partial G}{\partial P_{I_k}} \right). \tag{22}$$

【化 2 6】

加えて、 $M^T \pi^T = \pi^T$ であるので、

$$0 = \frac{\partial(M^T - I)\pi^T}{\partial P_{I_k}} \tag{23}$$

したがって、

$$\frac{\partial \pi^T}{\partial P_{I_k}} = -(M^T - I) + \frac{\partial(M^T - I)}{\partial P_{I_k}} \pi^T \tag{24}$$

$$\frac{\partial \pi}{\partial P_{I_k}} = -\pi \frac{\partial(M-I)}{\partial P_{I_k}} (M - I)^+ \tag{25}$$

【化 2 7】

最後に、式8で与えられた強化マルコフ連鎖の定義を使用して、

$$\frac{\partial M_{ik}}{\partial P_{I_k}} = \frac{M_{ij}}{\text{degout}(i) \cdot \sum_j M_{ij}^2} \tag{26}$$

$$\frac{\partial M_{ik}}{\partial P_{I_k}} = \frac{-M_{ij}}{\text{degout}(i) \cdot \sum_j M_{ij}^2} \quad (j \neq k, i \rightarrow j) \text{ の場合} \tag{27}$$

$$\frac{\partial M_{ik}}{\partial P_{I_k}} = 0 \quad (i \rightarrow j) \text{ 以外の場合} \tag{28}$$

式 1 4 ~ 式 2 8 の関係を式 1 3 の数式と共に使用して、攪乱指標に対する中心度値の感度の尺度を決定することができる。

【 0 1 1 1】

いくつかの実施形態では、結果の提示および解釈を改善するには、中心度値のフィルタリング、修正、またはフィルタリングと修正の両方を行うことが望ましい。具体的には、中心度値（図 5 の流れ図 5 0 0 のプロセスにより生成）は、ネットワークに対する攪乱の影響を視覚的に表すためのスペクトル変換ベクトルを使用して射影することができる。こうした状況で有用であるグラフ理論の 1 つのツールは、グラフ組合せラプラシアン（graph combinatorial Laplacian）である。組合せラプラシアンは、有向ネットワークの方向から独立しており、したがって、強化ランダムウォークに

10

20

30

40

50

関連して上に記載された因果関係を組み込むためにすぐに修正されることがない。したがって、ネットワークの因果律が除去される。具体的には、 G^0 が、 G の方向性を除去することによって（つまり、すべてのエッジを二方向性にするによって）定義される無向ネットワークを示すものとし、

【化 2 8】

$$L_{G^0}$$

を次式により定義されるグラフ組合せラプラシアンとする。

【化 2 9】

$$L_{G^0}(i, j) = \begin{cases} \text{deg}(i) & i = j \text{ の場合} \\ -1 & i \sim j \text{ の場合} \\ 0 & \text{そうでない場合} \end{cases} \quad (29)$$

10

具体的には、ノード i と j の間にエッジが存在する場合に数式 $i \sim j$ が満たされ、その結果、ラプラシアン

【化 3 0】

$$L_{G^0}$$

の行は、合計がゼロになる。ラプラシアン

【化 3 1】

20

$$L_{G^0}$$

は対称で正になり、したがって、そのスペクトルは正の実数になる。ネットワークの熱カーネルは、

【化 3 2】

$$\frac{\delta}{\delta t} f = -L_{G^0} f, \forall f \in l^2(V^0)$$

の基本解である。

【化 3 3】

$$e^{-t L_{G^0}}$$

30

によって表すことができる解の i 番目の行は、 i 、 ϕ_i におけるディラック熱源の拡散方程式の解を与える。加えて、 $g \in l^2(V^0)$ のスペクトル変換は、 g が、 m 個の実体を有するベクトルであり、次式により計算することができる。

【化 3 4】

$$F(g) = \sum_{i \in V^0} e^{-\lambda_i} \langle g | \phi_i \rangle \phi_i \quad (30)$$

ここで、 ϕ_i は

【化 3 5】

$$L_{G^0}$$

40

の固有ベクトルであり、 λ_i は対応する固有値である。具体的には、 $\langle g | \phi_i \rangle$ は g と ϕ_i の l^2 スカラー積である。一例では、 g は、

【化 3 6】

$$\frac{\langle g | \phi_i \rangle^2}{\|g\|_2^2}$$

が式 3 0 に使用されるような単位の大きさに正規化することができる。通常の慣習では、固有値を $0, \lambda_1, \lambda_2, \dots, \lambda_m$ に並び替える。いくつかの実施形態では、図 5 の流れ図 5 0 0 により計算された中心度値は、式 3 0 のスペクトル変換ベクトル上に射影す

50

ることができる。中心度値を射影し、限定された数のスペクトル変換ベクトルについての射影を表示するのみで、ノイズを軽減し、ネットワーク内の主要な経路を明らかにすることができる。このような射影は、多変量ネットワーク攪乱振幅 (NPA) 距離として使用して、実験的攪乱に対するネットワークモデルの応答を表すことができる。このような射影の例が図 13 および図 14 に提示されており、これらは、さまざまなノードに異なるパターンを使用して、2 つの最小非ゼロ固有値と関連付けられたスペクトル変換ベクトルの射影の値を示す。

【0112】

いくつかの実施形態では、ネットワークモデル内の複数のノードの中心度値を全体にわたって集約して、攪乱に対するネットワークモデルの応答を表すスカラー値を定義することが望ましい。上に記載されている多変量ネットワーク攪乱振幅 (NPA) 距離の代わりに、またはそれに加えて、スカラー値ネットワーク攪乱振幅 (NPA) 距離を使用して、実験的攪乱に対するネットワークモデルの応答を表すこともできる。上に記載されている中心度値は、任意の数の方法で組み合わせて、また任意の数の追加情報源と組み合わせて、スカラー値 NPA 距離を生成することができる。例えば、以下の手法のいずれか 1 つまたは複数を使用することができる。

10

【化 37】

$$1. \log_{10}(C_j): \sum_j \left| \log_{10} \left(\frac{C_j}{C_{SRW}(j)} \right) \right| \text{ の } l^2\text{-ノルム}$$

の l^2 ノルム (norm)。

20

2. 中心度値の \log_{10} のスペクトル変換のノルム (つまり、 \exp^{-j} で重みづけされたスペクトル変換ベクトル N_j 上への中心度比の射影の一次結合。中心度値を生成するためのトポロジーを使用することによって、またスペクトル変換ベクトルを生成するトポロジーも使用することによって、この手法では、非常に類似しているグローバル (スカラー値) スコアを有するが同じトポロジーのプロファイルは有さないことがある、2 つの攪乱を区別するための別レベルの細分性 (granularity) が得られる。

3. ランダム変数 $C = \max_j T_j$ によって定義される強化ランダムウォークのカバー時間。

【化 38】

$$\max_v E_v(C)$$

30

の正確な計算は計算的に困難であるが、上界は Matthew の定理によって次式により与えられる。

【化 39】

$$\sum_k^{n-1} \frac{1}{k}$$

この上界は、攪乱が全ネットワークに漸近的に伝搬する時間を表すので、NPA 距離を構築するのに使用することができる。

【0113】

40

細胞過程およびその攪乱の定量的解析について記述することは、疾患を理解する助けになる。生物学的過程間の非動力学的因果関係を記述するネットワークモデルが研究されてきた。このネットワークモデルでは、いくつかのノードが、そのノードによって記述される過程の下流標的に対応する 1 組の遺伝子と関連付けられる。モデルに含まれる挙動と、特定の試験における遺伝子発現レベルで観察される挙動との間の一致により、対応するノードの活性を定量化することができる。すなわち、ネットワークモデルは、短期間の分子の生物学的観察結果を疾患に関連する表現型端点と結びつけるのに役立つ。

【0114】

図 5 ~ 図 8 に関連して記載された中心度値技法は、ラットにホルムアルデヒドを曝露する試験に適用された。8 週齢の雄の F344 / Cr1BR ラットに、全身吸入によりホル

50

ムアルデヒドに曝露した。全身曝露は、0、0.7、2、6、10、および15 ppmの用量で実施した(1日当たり6時間、1週間当たり5日間)。動物は、曝露の開始後から1週間、4週間、および13週間で屠殺した。屠殺の後、鼻のレベルII領域からの組織を切除し、プロテアーゼの混合物で消化して上皮細胞を除去した。この鼻の切片から得られた上皮細胞は、主として移行上皮から成っており、気道上皮も一部あった。この上皮細胞に対して遺伝子発現マイクロアレイ分析を実施した。非疾患の哺乳動物肺細胞に対する攪乱の生物学的影響のシステムレベル評価を進めるために、肺に注目した細胞増殖の因果関係ネットワークが、WestraらのConstruction of a Computable Cell Proliferation Network Focused on Non-Diseased Lung Cells、BMC Systems Biology 2011、5:105で構築されており、この因果関係ネットワークは、正常肺細胞増殖を制御することにつながる多様な生物学の分野(細胞周期、成長因子、細胞相互作用、細胞内外シグナリング、およびエピジェネティクス)を包含し、合計848のノード(生物学的実体)および1597のエッジ(生物学的実体の間の関係)を含む。このネットワークは、肺および肺関連の細胞型の測定された細胞増殖端点と関連付けられた、4つの発表された遺伝子発現プロファイリングデータセットを使用して検証した。細胞周期制御に関与するコア機構(RB1、CDKN1A、およびMYC/MYC N)の活性の予測された変化は、複数のデータセットにわたって統計的に支持されており、それによって、システムの生物学データを使用してネットワーク全体の生物学的影響を評価するこの手法の全般的な適用可能性が強調される。図15に示された中心度結果は、ノードについての陰影の階調で示されている。具体的には、これらの結果は、いくつかのノード(例えば、Kaof(AktファミリーRn)に対応する大部分の明るい陰影のノード、WEE関連ノード、およびCdc2 P@Y15)が、強化されていないネットワークの領域を示す負のログ-中心度値を有することを示す。加えて、より明るい陰影の、負に影響するノード604(taof(E2F2)に対応)は、細胞増殖に負の影響を有する。別の例では、図15は、細胞増殖に関して正に影響するノード(taof(Myc)に対応)を示す。図15に示された結果は、taof(Myc)が細胞周期の制御に対して正の影響力のあるものである、ということを示す(例えば、相G1から相Sへの移行中に)。図15のノードの1つのサブセットが、測定可能な量の因果サイン(causal signature)のタイプと関連付けられているHYPを示す。「HYP」という名称は「仮説」に由来し、HYPが1組の予測を作成すると考えられ得るという事実を反映しており、このHYPにより、特定の生物学的過程の機構に関する洞察を得ることができる。具体的には、HYPは、1つまたは複数の測定可能実体(例えば、図15のノードの少なくとも一部)と、ある攪乱に応じてこれらが変化する方向(増加または減少)とに対応しうる。さらに、図16は、細胞増殖の強化における指数関数的な用量依存パターンを示し、これは、文献に記載されている結果と一致している。本明細書に記載されている技法を使用して、ネットワークの攪乱された領域が識別され、それにより、時間依存および用量依存の強化が明らかになるが、反対の兆候を有する領域もまた明らかになる。したがって、数千の下流制御された遺伝子の騒々しい挙動に隠されているシステム全体の応答の構造が、開示された手法によって捕捉され、それによって、生物学的ネットワークに対する外部攪乱の全体的な影響を記述する洞察力のある方法が、因果モデルに含まれる知識と遺伝子発現技術によって測定されたシステムの応答とを組み合わせることにより得られる。

【0115】

図9は、生物学的攪乱の影響を定量化する分散型のコンピュータ化されたシステム900のブロック図である。システム900のコンポーネントは、図1のシステム100内のものと同じであるが、システム100の配置構成は、それぞれのコンポーネントがネットワークインターフェース910を通じて通信するような構成をとる。そのような実装は、「クラウドコンピューティング」パラダイムなどの共通ネットワークリソースへのアクセスを共有することができるワイヤレス通信システムを含む複数の通信システム上での分散

10

20

30

40

50

コンピューティングに適している可能性がある。

【0116】

図10は、図1~10について説明されているプロセスを実行するための図1のシステム100または図9のシステム900のコンポーネントのうちの一つかなどのコンピューティングデバイスのブロック図である。SRPエンジン110、ネットワークモデリングエンジン112、ネットワークスコア化エンジン114、集約エンジン116、ならびに転帰データベース、攪乱データベース、および文献データベースを含むデータベースのうちの一つまたは複数を用意する、システム100のコンポーネントのそれぞれは、一つまたは複数のコンピューティングデバイス1000上に実装される。いくつかの態様において、複数の上記のコンポーネントおよびデータベースは、一つのコンピューティングデバイス1000内に収めることができる。いくつかの実装では、複数のコンピューティングデバイス1000にまたがって一つのコンポーネントおよび一つのデータベースを実装することができる。

10

【0117】

上記コンピューティングデバイス1000は、少なくとも一つの通信インターフェースユニット、入力/出力コントローラ1010、システムメモリー、および一つまたは複数のデータ記憶デバイスを備える。上記システムメモリーは、少なくとも一つのランダムアクセスメモリー(RAM1002)および少なくとも一つのリードオンリーメモリー(ROM1004)を備える。これらの要素はすべて、中央処理装置(CPU1006)と通信して、該コンピューティングデバイス1000の動作を円滑に行わせる。上記コンピューティングデバイス1000は、多くの異なる方法で構成される。例えば、上記コンピューティングデバイス1000は、従来のスタンドアロン型コンピュータであってもよいが、代替的に、コンピューティングデバイス1000の機能を複数のコンピュータシステムおよびアーキテクチャにまたがって分散させることもできる。上記コンピューティングデバイス1000は、モデリング、スコア化、および集約演算の一部または全部を実行するように構成される。図10では、上記コンピューティングデバイス1000は、ネットワークまたはローカルネットワークを介して、他のサーバもしくはシステムにリンクされる。

20

【0118】

上記コンピューティングデバイス1000は、分散型アーキテクチャで構成することができ、データベースおよびプロセッサは、別のユニットまたは場所に収納される。いくつかのこのようなユニットは、一次処理機能を実行し、最低限、汎用コントローラまたはプロセッサおよびシステムメモリーを含む。このような一態様では、これらのユニットのそれぞれは、通信インターフェースユニット1008を介して、他のサーバ、クライアントもしくはユーザコンピュータおよび他の関係するデバイスとの一次通信リンクとして働く通信ハブまたはポート(図示せず)に接続する。上記通信ハブまたはポートは、通信ルーターとしてもっぱら使用される、最小処理機能をそれ自体有することができる。さまざまな通信プロトコルが、システムの一部であってもよく、これは、限定はしないがイーサネット(登録商標)(Ethernet(登録商標))、SAP、SAS(商標)、ATP、BLUETOOTH(登録商標)、GSM(登録商標)、およびTCP/IPを含む。

30

40

【0119】

上記CPU1006は、一つまたは複数の従来のマイクロプロセッサなどのプロセッサ、および該CPU1006の操作負荷をオフロードする数値演算コプロセッサ(math co-processor)などの一つまたは複数の補助コプロセッサを備える。上記CPU1006は、上記通信インターフェースユニット1008および上記入力/出力コントローラ1010と通信し、これを通じて該CPU1006は他のサーバ、ユーザ端末、またはデバイスなどの他のデバイスと通信する。上記通信インターフェースユニット1008および上記入力/出力コントローラ1010は、例えば、他のプロセッサ、サーバ、またはクライアント端末と同時通信するための複数の通信チャネルを備えることができる。互いに通信するデバイスであっても、互いにひっきりなしに送信している必要はない

50

。それと反対に、そのようなデバイスは、必要に応じて互いに送信するだけでもよく、実際には大半の時間においてデータの交換を差し控えることができ、該デバイス間の通信リンクを確立するために実行するのにいくつかのステップを必要とするものとしてよい。

【 0 1 2 0 】

上記 CPU 1 0 0 6 は、上記データ記憶デバイスとも通信する。上記データ記憶デバイスとして、磁気メモリー、光メモリー、または半導体メモリーの適切な組み合わせを含み得、例えば、RAM 1 0 0 2、ROM 1 0 0 4、フラッシュドライブ、コンパクトディスクなどの光ディスク、またはハードディスクもしくはドライブが挙げられる。上記 CPU 1 0 0 6 および上記データ記憶デバイスはそれぞれ、例えば、単一のコンピュータまたは他のコンピューティングデバイス内に丸ごと配置されるか、または USB ポート、シリアルポートケーブル、同軸ケーブル、イーサネット（登録商標）型ケーブル、電話回線、無線周波トランシーバ、または他の類似のワイヤレスもしくは有線媒体または上記のもの

10

【 0 1 2 1 】

上記データ記憶デバイスは、例えば、(i) 上記コンピューティングデバイス 1 0 0 0 用のオペレーティングシステム 1 0 1 2、(i i) 本明細書に記載されているシステムおよび方法により、また特に上記 CPU 1 0 0 6 に関して詳しく記載されているプロセスにより、該 CPU 1 0 0 6 に指示するように適合された 1 つまたは複数のアプリケーション 1 0 1 4（例えば、コンピュータプログラムコードまたはコンピュータプログラム製品）、または (i i i) 上記プログラムが必要とする情報を記憶するために利用されうる情報を記憶するように適合されたデータベース（1 つまたは複数） 1 0 1 6 を記憶することができる。いくつかの態様では、上記データベース（1 つまたは複数）として、実験データを記憶するデータベース、および公開文献モデルが挙げられる。

20

【 0 1 2 2 】

上記オペレーティングシステム 1 0 1 2 およびアプリケーション 1 0 1 4 は、例えば、圧縮形式、非コンパイル形式、および暗号化形式で記憶され、コンピュータプログラムコードを含むことができる。上記プログラムの命令は、上記 ROM 1 0 0 4 または上記 RAM 1 0 0 2 などの、データ記憶デバイス以外のコンピュータ可読媒体から上記プロセッサのメインメモリーへと読み込むことができる。上記プログラムにおける命令のシーケンスの実行により上記 CPU 1 0 0 6 が本明細書に記載されているプロセスステップを実行するが、ハード配線回路を、本開示のプロセスの実装のためのソフトウェア命令の代わりに、または該ソフトウェア命令と組み合わせて使用することができる。したがって、記載されているシステムおよび方法は、ハードウェアとソフトウェアとの特定の組み合わせに限定されない。

30

【 0 1 2 3 】

本明細書に記載されているようなモデリング、スコア化、および集約に関して 1 つまたは複数の機能を実行するのに適したコンピュータプログラムコードが提供されうる。上記プログラムは、オペレーティングシステム 1 0 1 2、データベース管理システム、および上記プロセッサが上記入力/出力コントローラ 1 0 1 0 を介してコンピュータ周辺デバイス（例えば、ビデオディスプレイ、キーボード、コンピュータマウスなど）とインターフェースすることを可能にする「デバイスドライバ」などのプログラム要素を含むことができる。

40

【 0 1 2 4 】

本明細書で使用されているような「コンピュータ可読媒体」という用語は、実行のため命令を上記コンピューティングデバイス 1 0 0 0（または本明細書に記載されているデバイスの任意の他のプロセッサ）のプロセッサに与えるか、または与えることに関わる任意の非一時的媒体を指す。このような媒体は、限定はしないが、不揮発性媒体および揮発性

50

媒体を含む、多くの形態をとりうる。不揮発性媒体としては、例えば、光ディスク、磁気ディスク、もしくは光磁気ディスク、またはフラッシュメモリなどの集積回路メモリが挙げられる。揮発性媒体としては、典型的にはメインメモリを構成するダイナミックランダムアクセスメモリ（DRAM）を含む。コンピュータ可読媒体の一般的な形態としては、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVD、任意の他の光媒体、パンチカード、紙テープ、穴の形状を有する任意の他の物理的媒体、RAM、PROM、EPROMまたはEEPROM（電子的に消去可能なプログラム可能リードオンリーメモリ）、FLASH-EEPROM、任意の他のメモリチップもしくはカートリッジ、またはコンピュータが読み取ることができる任意の他の非一時的媒体が挙げられる。

10

【0125】

さまざまな形態のコンピュータ可読媒体が、実行のため1つまたは複数の命令の1つまたは複数のシーケンスを上記CPU1006（または本明細書に記載されているデバイスの任意の他のプロセッサ）に伝えることに関与しうる。例えば、上記命令は、最初に、リモートコンピュータ（図示せず）の磁気ディスクで伝えることができる。上記リモートコンピュータは、命令をそのリモートコンピュータのダイナミックメモリにロードし、モデムを使用してイーサネット（登録商標）接続、ケーブル線、さらには電話回線を介して該命令を送ることができる。コンピューティングデバイス1000（例えば、サーバ）に対してローカルの通信デバイスは、各通信回線上でデータを受け取り、該データを上記プロセッサのシステムバスに出すことができる。上記システムバスは、データをメインメモリに伝え、上記プロセッサはそのメインメモリから命令を取り出して実行する。メインメモリに入った命令は、必要に応じて、上記プロセッサによる実行前または実行後にメモリに記憶することができる。それに加えて、命令は、通信ポートを介して、電気信号、電磁気信号、または光信号として受け取ることができ、これらはさまざまな種類の情報を伝えるワイヤレス通信またはデータストリームの形態の例である。

20

【0126】

さらなる一態様では、生物系のネットワークモデル内のノードについての距離を決定するためのコンピュータシステムが提供され、このコンピュータシステムは、作用物質に対する生物系の応答に対応する処置データの組を受け取るように構成または適合された第1のプロセッサであって、生物系が複数の生物学的実体を含み、それぞれの生物学的実体が、各生物学的実体のうちの他の少なくとも1つと相互作用する第1のプロセッサと、作用物質に曝露していない生物系に対応するコントロールデータの組を受け取るように構成または適合された第2のプロセッサと、計算因果関係ネットワークモデルを提供するように構成または適合された第3のプロセッサであって、この計算因果関係ネットワークモデルが生物系を表すと共に、生物学的実体を表すノードおよび生物学的実体間の関係を表すエッジを含み、エッジが、対応する第1のノードに対応する第2のノードに接続する第3のプロセッサと、ネットワークモデルに少なくとも一部は基づいて、ノードのサブセットの攪乱指標を計算するように構成または適合された第4のプロセッサであって、攪乱指標が、対応するノードにおける処置データとコントロールデータの間の差を表し、また対応するノードの活性が攪乱から影響を受ける程度を表す第4のプロセッサと、攪乱指標に少なくとも一部は基づいて、エッジの遷移確率を計算するように構成または適合された第5のプロセッサであって、エッジの遷移確率が、対応する第1のノードから対応する第2のノードへの遷移の尤度を表す第5のプロセッサと、遷移確率に少なくとも一部は基づいて、ノードの中心度値を生成するように構成または適合された第6のプロセッサであって、中心度値がネットワークモデル内の対応するノードの相対的重要度を表す第6のプロセッサ、とを備える。

30

40

【0127】

さらなる一態様では、第1の処置データの組を受け取るように構成または適合された第1のプロセッサと、第2の処置データの組を受け取るように構成または適合された第2のプロセッサと、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジ

50

を含む計算因果関係ネットワークモデルを提供するように構成または適合された第3のプロセッサと、ネットワークモデルに少なくとも一部は基づいてノードのサブセットの攪乱指標を計算するように構成または適合された第4のプロセッサであって、攪乱指標が、対応するノードにおける第1と第2の処置データの間の差を表す第4のプロセッサと、攪乱指標に少なくとも一部は基づいて対応するノードの中心度値を生成するように構成または適合された第5のプロセッサであって、中心度値がネットワークモデル内の対応するノードの相対的重要度を表す第5のプロセッサと、第2のノードの攪乱指標に関する第1のノードの中心度値の偏導関数を計算するように構成または適合された第6のプロセッサであって、偏導関数がネットワークモデルのトポロジー感度尺度を表す第6のプロセッサとを備える、コンピュータシステムが提供される。

10

【0128】

さらなる一態様では、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む計算ネットワークモデルを提供するように構成または適合された第1のプロセッサと、ネットワークモデルに少なくとも一部は基づいて、対応するノードの中心度値を生成するように構成または適合された第2のプロセッサであって、中心度値がネットワークモデル内の対応するノードの相対的重要度を表す第2のプロセッサと、ネットワークモデルに対する攪乱の影響を表すためのスペクトル変換ベクトル上への中心度値の射影を計算するように構成または適合された第3のプロセッサとを備える、コンピュータシステムが提供される。

【0129】

20

さらなる一態様では、生物系の攪乱を定量化するコンピュータシステムが提供され、このコンピュータシステムは、生物学的実体を表すノード、および生物学的実体間の関係を表すエッジを含む計算因果関係ネットワークモデルを提供するように構成または適合された第1のプロセッサと、ネットワークモデルに少なくとも一部は基づいて、対応するノードの中心度値を生成するように構成または適合された第2のプロセッサであって、中心度値がネットワークモデル内の対応するノードの相対的重要度を表す第2のプロセッサと、中心度値を集約して、生物系の攪乱を表すネットワークモデルのスコアを生成するように構成または適合された第3のプロセッサとを備える。

【0130】

さらなる一態様では、本明細書に記載されている方法を実施するように適合されたプログラムコードを含むコンピュータプログラム製品が提供される。

30

【0131】

さらなる一態様では、コンピュータプログラム製品を含むコンピュータまたはコンピュータ可読媒体またはデバイスが提供される。

【0132】

本開示の実装は、特定の例を参照しつつ特に示され記載されているが、当業者であれば、添付の特許請求の範囲に定められているように本開示の精神および範囲から逸脱することなく形態および細部にさまざまな変更を加えられることを理解するはずである。そこで、本開示の範囲は、添付の特許請求の範囲によって示され、したがって、該特許請求の範囲の等価性の意味および範囲内にあるすべての変更は、包含されることが意図されている。上記明細書で述べられているすべての刊行物は、参照により本明細書に組み込まれる。

40

【 図 1 】

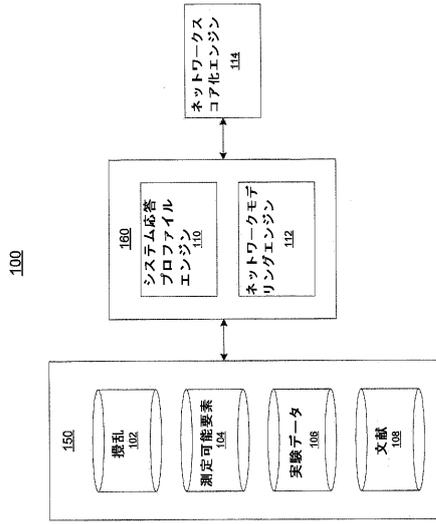


FIG. 1

【 図 2 】

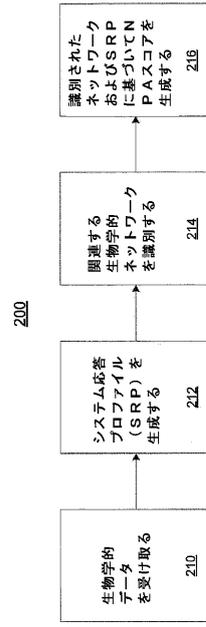


FIG. 2

【 図 3 】

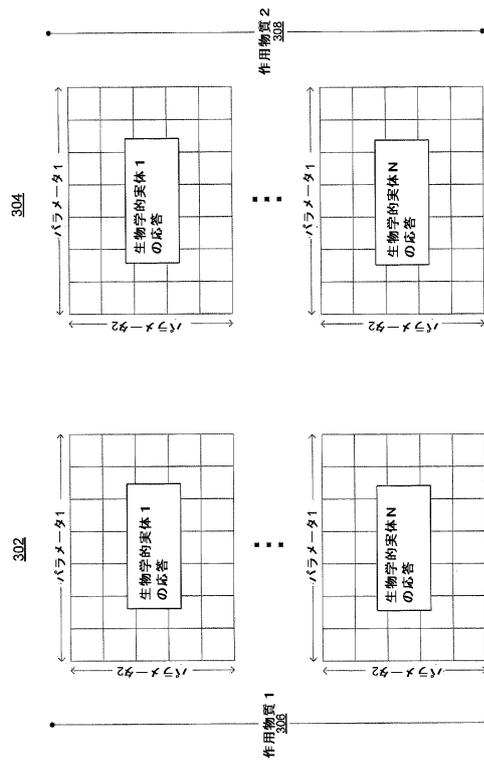


FIG. 3

【 図 4 A 】

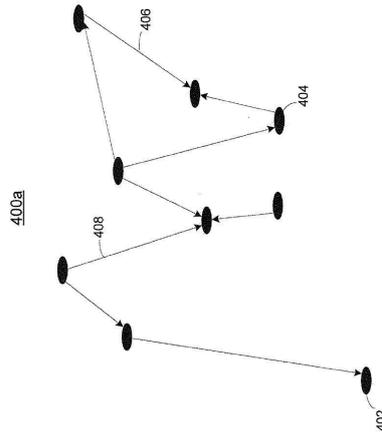
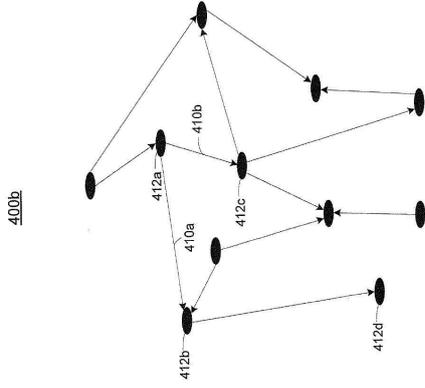


FIG. 4A

【 図 4 B 】



【 図 5 】

FIG. 4B

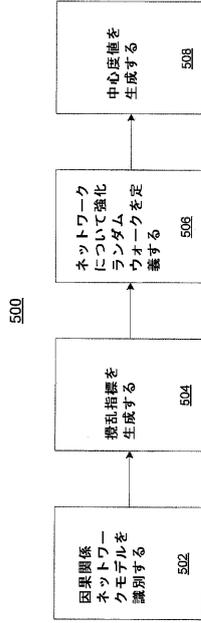


FIG. 5

【 図 6 】

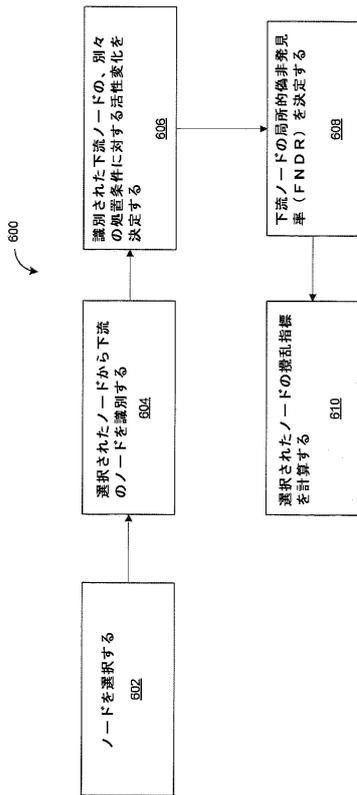


FIG. 6

【 図 7 】

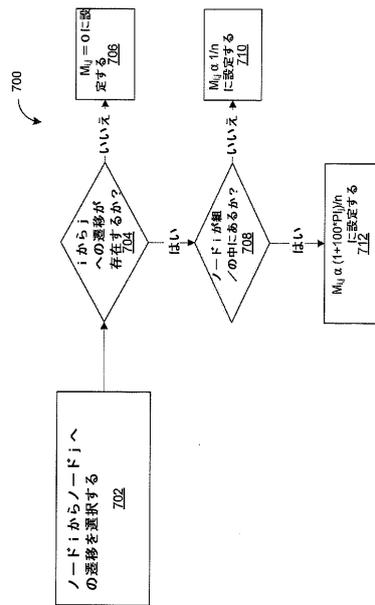


FIG. 7

【 図 8 】

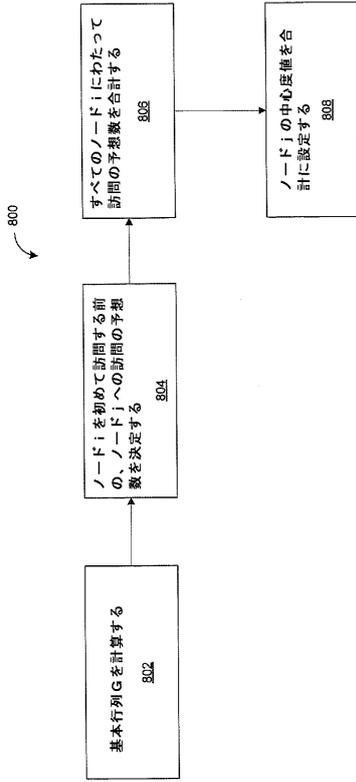


FIG. 8

【 図 9 】

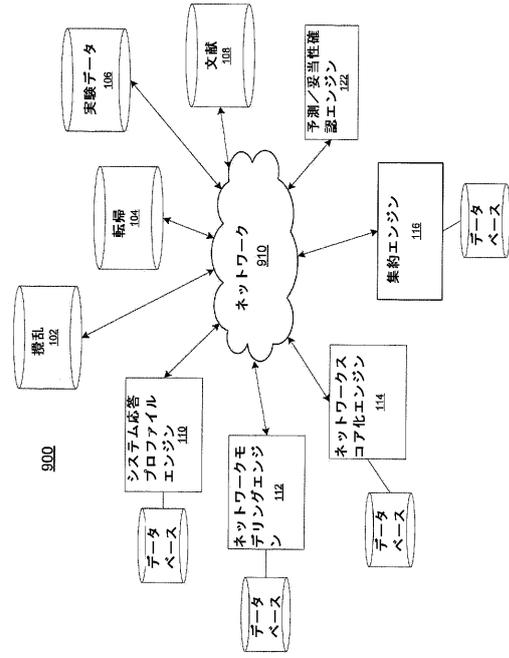


FIG. 9

【 図 10 】

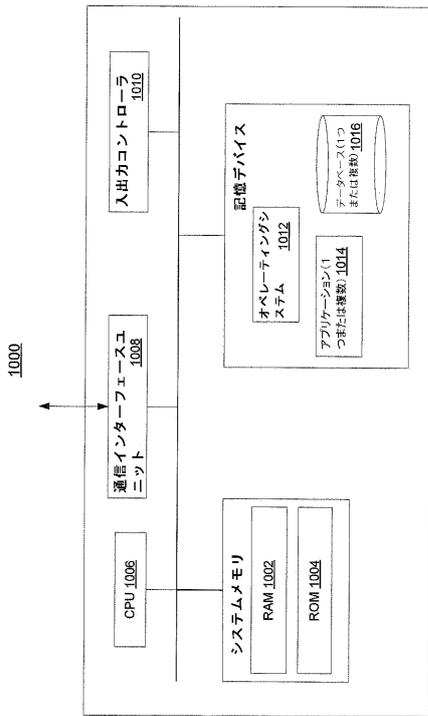


FIG. 10

【 図 11 】

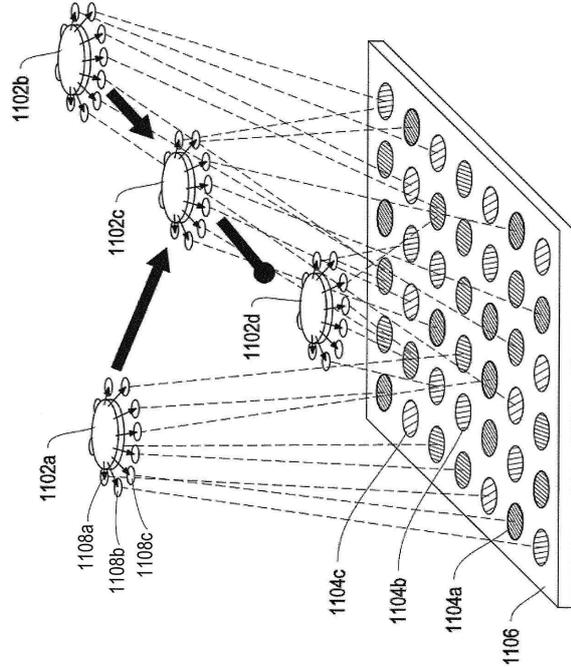


FIG. 11

【 図 1 2 】

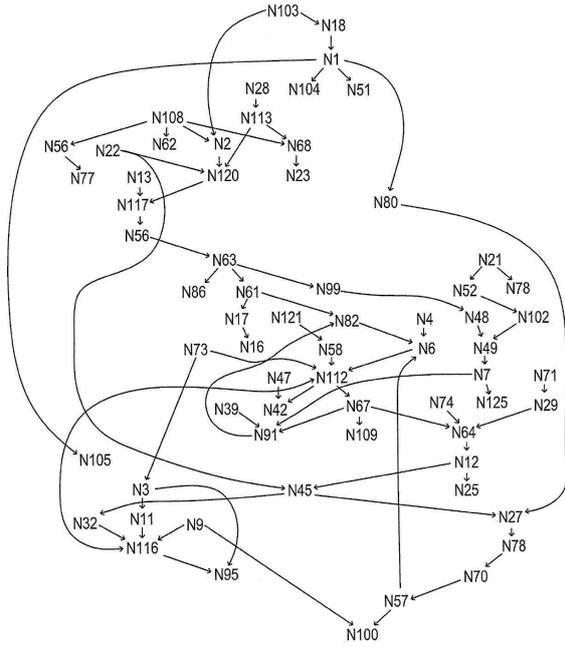


FIG. 12

【 図 1 3 】

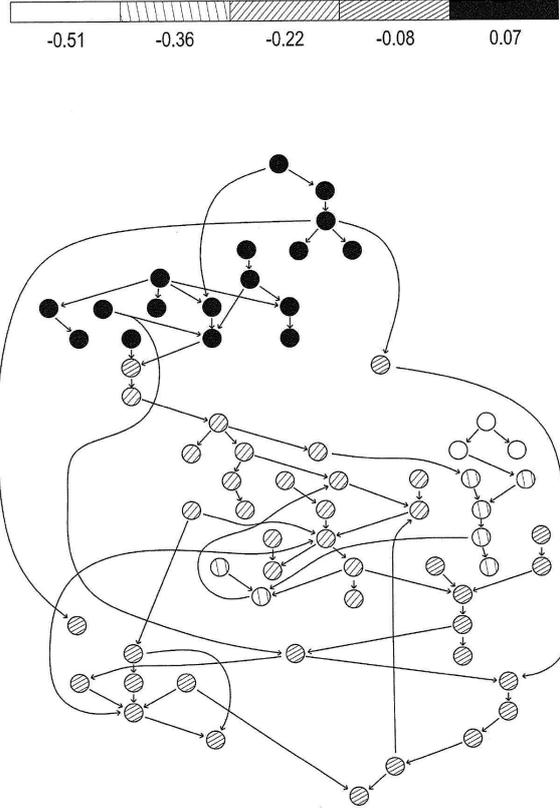


FIG. 13

【 図 1 4 】

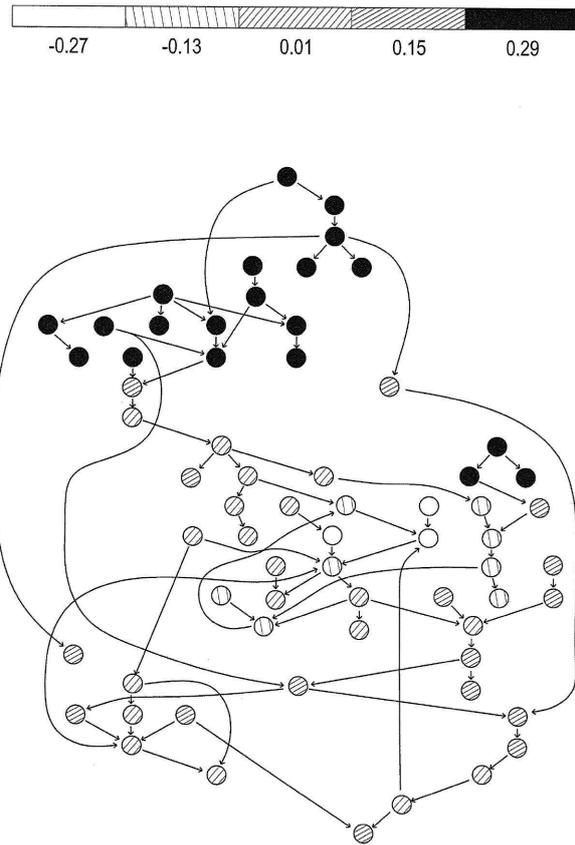


FIG. 14

【 図 1 5 - 1 】

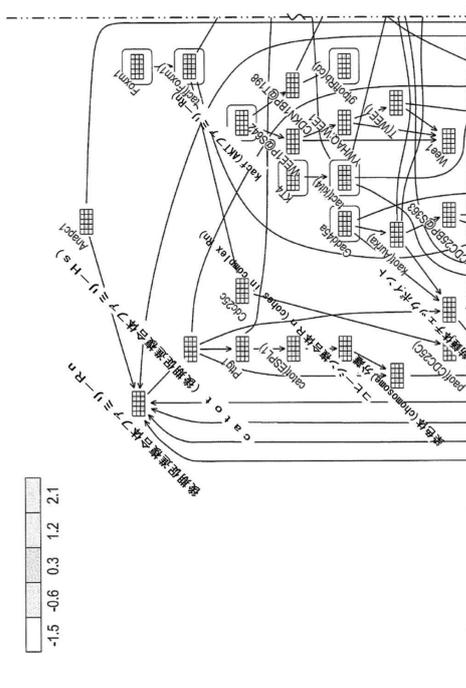
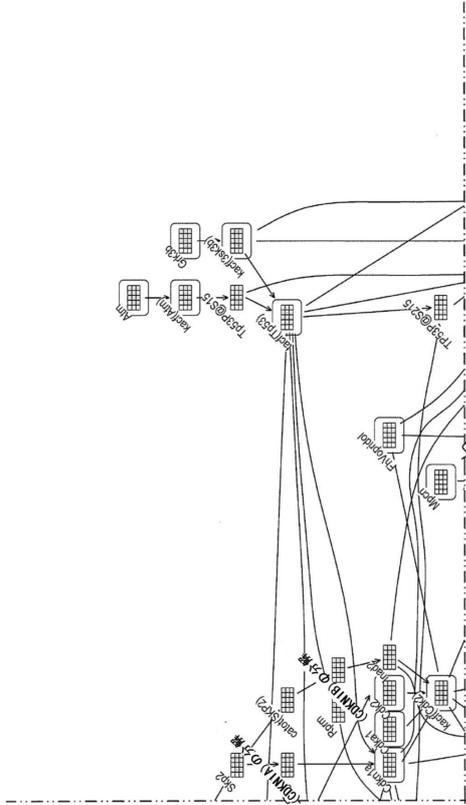


FIG. 15 (part 1)

【図 15 - 2】



【図 15 - 3】

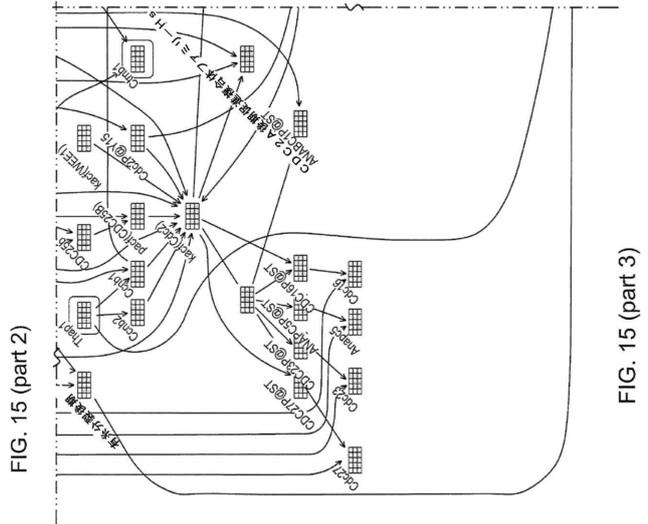


FIG. 15 (part 2)

FIG. 15 (part 3)

【図 15 - 4】

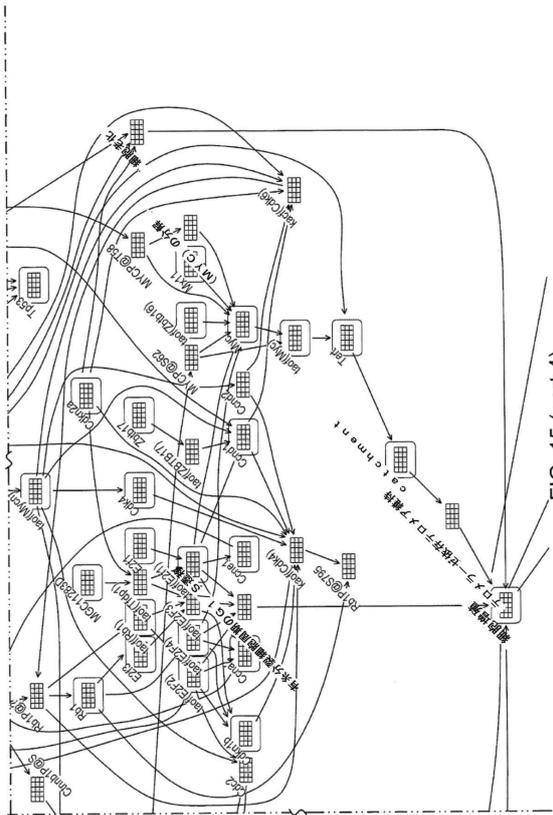


FIG. 15 (part 4)

【図 16】

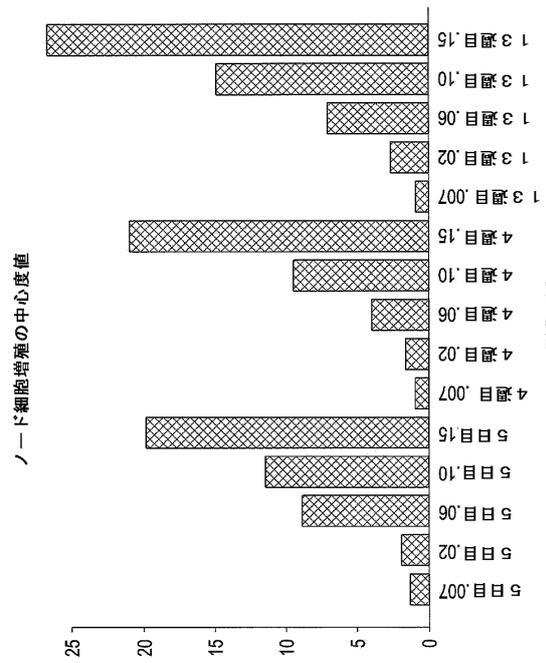


FIG. 16

フロントページの続き

(72)発明者 フロリアン マルティン

スイス国 ツェーハー - 2 0 3 4 プスー, シュマン ドゥ ロレー 1

(72)発明者 アラン セーヴェル

スイス国 ツェーハー - 3 9 7 7 グランジュ, リュ ドゥ ラ レシュ 3

審査官 塩田 徳彦

(56)参考文献 REVERSE CAUSAL REASONING METHODS - WHITE PAPER, INTERNET CITATION [ONLINE], 2 0 1 1年
2月 4日, P1-26, <http://www.selventa.com/sites/g/files/g813951/f/RCR%20Whitepaper.pdf>, U R L , http://www.selventa.com/attachments/white_papers/reverse-causal-reasoning.pdf

(58)調査した分野(Int.Cl., D B名)

G 0 6 F 1 9 / 1 0 - 1 9 / 2 8

C 1 2 M 1 / 0 0

C 1 2 Q 1 / 6 8