

文書ストリームにおける語のバーストと共起ネットワークにおけるクラスタ構造の関係について

佐藤進也^{†1} 福田健介^{†2}
菅原俊治^{†3} 栗原聡^{†4}

文書に現れる語をノードとし、出現位置が近接しているものどうしをリンクで結び付けることにより得られる共起ネットワークでは、意味的関連性を有する語どうしがクラスタ構造と呼ばれる稠密な相互のつながりを形成している。本論文では、時間経過にともない文書が生成されていく、いわゆる文書ストリームから共起ネットワークを構成し、そこでクラスタ構造が生成される様子を調べた。その結果、共起ネットワークを（相対的に）古い語彙からなる部分と新しい語彙からなる部分に分けたとき、後者において、クラスタの出現という構造上の変化が、実社会の出来事などに起因する語の出現頻度の増大（バースト）に関連していることが明らかになった。

On the Relationship between Word Bursts in Document Streams and Clusters in Lexical Co-occurrence Networks

SHIN-YA SATO,^{†1} KENSUKE FUKUDA,^{†2} TOSHIHARU SUGAWARA^{†3}
and SATOSHI KURIHARA^{†4}

In word co-occurrence networks, where two words appearing close to each other in documents are connected by a link, a group of relevant words forms a densely connected subnetwork called a cluster. In this paper, we analyze the process of emergence of the structure in co-occurrence networks generated from document streams. The analysis reveals that, if we restrict our scope to the subnetwork mostly consisting of (relatively) new words, we can associate emergence of the structure with the increase of word occurrence rate (bursts) that arises from real world events.

1. はじめに

共起ネットワークとは、文書に現れる語をノードとし、出現位置が近接しているものどうしをリンクで結び付けることにより得られるネットワークである。

一般に、文書中の出現位置が近接している語どうしは同じ文脈を共有していると考えられる。よって、共起ネットワーク上で互いに近くに位置する語どうしには多くの場合意味的な関連性が認められる。図1は、1995年の新聞記事をもとに構成した共起ネット

ワークの中で“ライフライン”に隣接する語どうしのつながり示したものである。中央に、“電気”、“ガス”、“水道”、“電話”といったライフラインの設備が互いに密につながっており、上方には“ライフライン”という語が使われる文脈を示す震災関係の語が現れている。

さらに、共起ネットワークでは、文脈を共有し関連性を有する語は相互に結び付けられるので局所的に稠密な構造（クラスタ構造）を形成する。その一方で、関連性の低い語どうしの結び付きは疎になる。この性質を利用し、ネットワークの構造解析による語の曖昧性（多義性）の解消や語どうしの関連性の推定が試みられてきた（次章参照）。

本論文では、この特徴的構造の応用から生成原理の解明に視点を移す。具体的には、時間経過にともない文書が生成されていく、いわゆる文書ストリームにおいて、この構造が生成される様子を調べる。そして、共起ネットワークを古い部分と新しい部分に分けたとき、後者において、クラスタ構造の出現という構造上

^{†1} NTT 未来ねっと研究所

NTT Network Innovation Laboratories

^{†2} 国立情報学研究所

National Institute of Informatics

^{†3} 早稲田大学

Waseda University

^{†4} 大阪大学産業科学研究所

The Institute of Science and Industrial Research, Osaka University

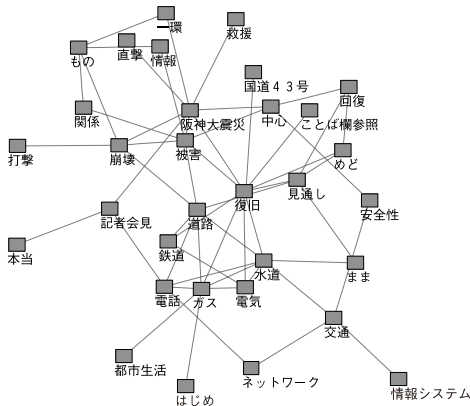


図1 共起ネットワークの例

Fig. 1 Example of a word co-occurrence network.

の変化が、実社会における出来事の発生などにもなう語の出現頻度の増大（バースト）に関連していることを明らかにする。この事実を、新しい語彙からなる部分ネットワーク中のクラスタ構造は出来事の発生にもなうて生み出される新しい文脈を表している、と解釈することができる。

また、これは、文書ストリーム、あるいは一般に文書集合から出来事を検出するという応用目的の観点からも有用な知見である。すなわち、本研究は、従来バーストという時間軸上の変化でとらえられてきた出来事が、クラスタ構造の出現という構造上の変化をとらえることでも検出可能であるという新しい知見を示すものである。

以下、本論文での議論を次のようにすすめる。まず、本論に先立って、本論文での議論に関わる基本的用語の記法および用法を2章で明確にしておく。その後、3章で共起ネットワークの解析に関する先行研究について述べ、本研究の位置づけを明らかにする。次いで、本研究で解析する共起ネットワークの構成方法を4章で述べる。5章では、近年において、日本の社会に最も大きなインパクトを与えた事象の1つである阪神淡路大震災を題材とし、出来事の発生にもなう共起ネットワークの構造変化を調べる。そして、共起ネットワークの部分ネットワークにおいて、出来事に由来してクラスタ構造が現れることをいくつかの例により確認する。さらに、出来事とクラスタ構造の関連性を詳しく調べるため、6章では出来事をそれに関係する語のバーストに対応付け、バーストとクラスタ構造の関係を定量的に解析する。この結果をふまえて、7章では、社会的に認知されている大きな出来事を題材にして、出来事、バースト、クラスタ構造の対応関係を確認する。そして、8章では本手法の性質と適用

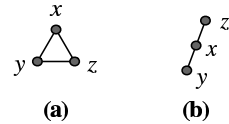


図2 つながりの稠密さの違い

Fig. 2 Difference in density of interconnections.

限界を整理する。

2. 用語の定義

2.1 基本的な記法と用語

ネットワーク X は、ノードと、ノードどうしを結び付けているリンクから構成される。つまり、 X は、形式的には、ノードの集合 N_X とリンクの集合 L_X の組 (N_X, L_X) で表すことができる。ノードの数 $|N_X|$ をネットワーク大きさと呼び $|X|$ で表す。また、ネットワーク X, Y の共通部分ネットワーク $Z = (N_Z, L_Z)$ を $X \cap Y$ で表し、次のように定義する：

$$N_Z = N_X \cap N_Y$$

$$L_Z = \{l | l \in L_X \cup L_Y, V(l) \subset N_Z\}$$

ここで、 $V(l)$ はリンク l が結び付けている2つのノードからなる集合である。

ノード x の次数 $k(x)$ とは、 x と他のノードとのつながりの数であり、たとえば次のように定義することができる：

$$k(x) = |\{l | x \in V(l)\}|$$

2.2 クラスタ構造とクラスタ性

いま、図2のように、ノード x が他のノード y, z とつながっている2つの状況を考える。(a) はすべてのノードが互いにつながっており、(b) よりも密な相互関係が形成されていると考えられる。この違いは、クラスタ係数²¹⁾により定量的に把握することができる。

ノード x のクラスタ係数は、一般に、次のように定義される。 x が i 個のノードとリンクでつながっているとき、それらのノードの間には iC_2 (i 個の要素から2個を選ぶ組合せの数)本のリンクが存在する。実際に存在するリンクの数が j 本であるとき、 x のクラスタ係数は $j/iC_2 = 2j/(i(i-1))$ で与えられる。図2のそれぞれについて x のクラスタ計算すると、(a) では1であり、(b) では0となる。

本論文では、クラスタ係数の向上につながる構造を総称してクラスタ構造と呼ぶ。その最も基本的なものが、図2(a)に示した、3つのノードが互いにつながりあっている三角形である。

また、本論文では、ノード x のクラスタ係数が正であるとき、別ないい方をすれば、ノードの近傍がクラスタ構造を含むとき、ノード(の近傍)がクラスタ

性を持つという．さらに，クラスタ係数が大きな値を示す状況を，クラスタ性が高いという．

なお，本論文では，ノード x 自身および x とリンクでつながっているノードと，それらの間のリンクからなるネットワークを語 x の近傍と呼び， $\Gamma(x)$ で表すことにする．また， $\Gamma(x)$ から x を除去したネットワークを $\Gamma_*(x)$ と書く．

2.3 バーストとバースト性

構造的な稠密さをクラスタ係数により定量的に把握し，その値が正となるか否かをクラスタ性の有無に対応させたのと同様に，本論文では，バーストを測る（検出する）ための尺度であるバースト度を導入し（6.1 節）その値によりバースト性を定義する．すなわち，語 x のバースト度が正値か否かでバースト発生の有無を判定し，（ある期間内に）発生が認められる場合に x はバースト性を持つということにする．

3. 関連研究

1 章でも述べたように，共起ネットワークでは，関連する語が互いに結び付けられ稠密な構造を形成する一方で，関連性の低い語どうしの結び付きは疎になる．この性質を利用し，つながりの疎密に着目してネットワークの構造を解析することにより，語の意味把握や曖昧性（多義性）を解消する手法が考えられてきた^{3),13),19),20)}．また，意味の同異性，関連性が反映されている共起ネットワークの構造を可視化することにより，ユーザによる情報（文書集合）の理解を支援するという手法も提案されている¹⁸⁾．さらに，ネットワークの構造的な特徴から，文書の主題を示すような重要語を抽出するというアプローチもある^{11),14)}．このように，共起ネットワークは，ことばの意味理解などへの応用を目的として解析されてきた．

一方，近年では，複雑ネットワーク科学¹⁰⁾ のアプローチが注目されている．これは，多数の要素からなる複雑な対象の解析手法として，要素間の関係をネットワークで表現し，その静的な特徴だけでなく，成長過程などの動的な特徴（ダイナミクス）を調べるといったものであり，様々な分野で適用が試みられている¹²⁾．言語に関しても，この手法を採り入れ，その基本原理を解明しようとする取り組みがある^{2),4),15),17)}．たとえば，文献 2) では，語彙の成長とは独立な，言語の核となる語（kernel lexicon）を，共起ネットワー

クの次数分布の違いにより識別できることを示した．

本研究もまた，語どうしの共起という関係を表すネットワークのダイナミクスに着目しているという意味で，複雑ネットワーク科学の視点からの言語解析の範疇に属すると考えられる．その一方で，本研究は，ことばの意味の識別という，従来からの共起ネットワーク解析の目的を継承している．つまり，本研究は，従来よりある意味把握の要求に対する複雑ネットワークの視点に立ったアプローチとして特徴付けられる．

4. 共起ネットワークの生成方法

4.1 ノードとなる語の選出とリンク生成の方法

本論文では，文書ストリームとして新聞記事⁹⁾ を使い，そこに出現する語で共起ネットワークを生成する．その具体的な方法は次のとおりである．まず，記事を構成する文ごとに形態素解析し，名詞あるいは未知語を抜き出す．その中で連続して出現する語は，それらをつなげて 1 つの語と見なす．このようにして得られた語をノードとし，出現順に前後の関係のあるものをリンクで結ぶ．たとえば，「阪神大震災は神戸市などの一部で震度 7 を記録した」という文からは「阪神」，“大震災”，“神戸”，“市”，“一部”，“震度”，“7”，“記録”という名詞が抜き出される．そして，“阪神”と“大震災”，“神戸”と“市”，“震度”と“7”がそれぞれ連続して現れているので，この文から生成される共起ネットワークは以下ようになる：



4.2 古い語彙と新しい語彙

ここで，共起ネットワークを構成する語を，次のようにして，古いものと新しいものとに二分することを考える．まず，（相対）時刻 $t = 0$ に対応する基準日 T_0 を選ぶ．さらに，負の整数値をとるパラメータ τ_{pre} を決め，記事が書かれた時刻（日） t について， $\tau_{pre} < t \leq 0$ の記事に現れる語と， $t > 0$ である記事にはじめて出現する語を区別する．本論文では，前者を「古い」語彙，後者を「新しい」語彙と呼ぶことにする．図 3 は，時間の経過とともに順次記事が生産され，そこに語（a～e）が出現している様子を模式的に示したものである．これを例にとると，語 a, b, c は古い語彙に属し，d, e は新しい語彙に属する．

x の除去にともない， x とその他のノードをつないでいるリンクも除かれる．
人々により構成される社会組織や神経細胞のつながり，インターネットなど．

ここでの狙いは，個々のことばが何を意味しているのかを明らかにすることではなく，「ネットワーク上で密につながったものどうしを意味的なまとまりがあるものとしてとらえその他のものと区別する」というように，意味の相違を把握することである．

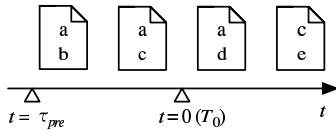


図3 時間経過にともなう記事と語の出現の模式図

Fig. 3 Schematic example of a document stream and words appearing in the documents.

表1 “ボランティア”の近傍に属する語彙の例

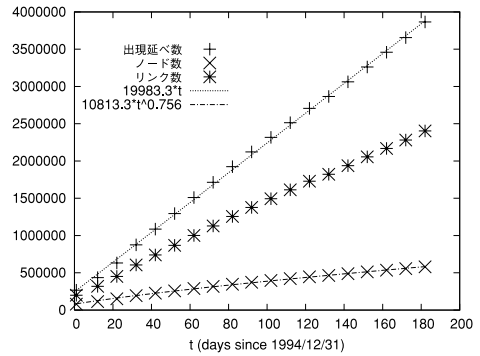
Table 1 Vocabulary found in the neighborhood of “ボランティア”.

古い語彙	新しい語彙
欧米, 行政, 思い入れ, 試み, これ, 手伝い, 人, 対象, 団体, 地元, 動き, 日本赤十字社, 派遣, 募集, 役割	義援金, 阪神大震災, 炊き出し, 西宮市立中央体育館, 青空給食, 特別立法, 非難所生活, 復旧活動

一般に、高い確率で出現する語は古い語彙に属している可能性が高い。そして、そのような語彙には、上記の例でいえば“一部”のような、文脈に非依存であり、共起ネットワークにおいて多くの語と結び付きを持っている語が含まれる。つまり、語彙を二分することには、新しい文脈で用いられている語を古い文脈のものとは区別するという意図があるのだが、さらに新しい語彙と文脈非依存な語彙を区別するという意味合いがある。

例として、 T_0 を 1994 年 12 月 31 日、 $\tau_{pre} = -14$ としたとき、“ボランティア”の近傍に属するそれぞれの語彙を表 1 に示す。

以降、特に断らない限り、 T_0 は 1994 年 12 月 31 日、 τ_{pre} の値は -14 に固定し、共起ネットワークといえば時間区間 $(\tau_{pre}, t]$ の間に書かれた記事から生成されるネットワーク $\mathcal{K}(t)$ を指すものとする。 t を $\mathcal{K}(t)$ の作成時刻と呼ぶ。 $\mathcal{K}(0)$ は古い語彙のみからなるネットワークである。これを「土台」として、 $\mathcal{K}(t)$ がどのように成長していくかを調べるのが本論文の主題である。なお、 $t > 0$ に書かれた記事における古い語彙に属する語の出現は、 $\mathcal{K}(t)$ のノード数の増加には寄与しないが、リンク数の増加をもたらす可能性があることに注意されたい。図 4 は、記事に出現した語の延べ数、 $\mathcal{K}(t)$ のノード数、リンク数の変化を示したものである。語の延べ数はほぼ線形な増加を示しているのに対し、ノード数 (= 語彙数) の増加の割合は t の増加にともない鈍っているのが分かる。

図4 共起ネットワーク $\mathcal{K}(t)$ の時間発展Fig. 4 Time evolution of the co-occurrence network $\mathcal{K}(t)$.

5. 「出来事」が生み出すネットワークの局所的な変化

図 4 はネットワークのマクロな時間発展の様子を示したものである。これを見る限り、急激な変化や増減の繰返しなどの複雑な変動はなく、単調に成長しているように見える。しかし、ミクロなレベルでは複雑な振舞いが観測される。特に、大きな出来事が発生したときには、それに関連する語の近傍でも大きな変化が現れる。本章では、1995 年 1 月 17 日 ($t = 17$) に発生した阪神淡路大震災を題材とし、この出来事と関連性があり、後に示すようにこの出来事に降急速に出現が増えた語である“ボランティア”を選び、その近傍の変化を調べる。以降、 v という記号で語“ボランティア”を表すことにする。

5.1 特徴量による近傍の状況の把握

図 5 は、 $\mathcal{K}(t)$ における v の次数の変化を示したグラフである。比較のために、 $t = 0$ の時点で v と次数が等しい“歌手”と“リーダーシップ”の次数の変化をあわせて表示してある。この 3 語の中で、 v の次数だけが、震災発生直後から急激に増加していることが見てとれる。

この変化をさらに詳しく調べた結果を図 6 に示す。これは、 v の近傍の状況を、下記 4 種類の特徴量によりとらえたものである。それぞれの縦軸のスケールは、グラフが図に収まるように適宜調整してある。

(i) 日ごとの出現頻度

t 日目に書かれた記事の中に v が出現した回数。

(ii) クラスタ係数

日本においては、“ボランティア”には、それを趣味とするようなある特定の人々が行うものというニュアンスがあったが、1995 年の阪神淡路大震災で（それまでのボランティア経験の有無にかかわらず）様々な人々が全国から駆けつけた。これにより、1995 年は「ボランティア元年」と呼ばれている。

文書数 n とそこから得られる語彙の数 v には Heaps の経験則と呼ばれる次の関係が成り立つことが知られている： $v = Kn^\beta$ 。

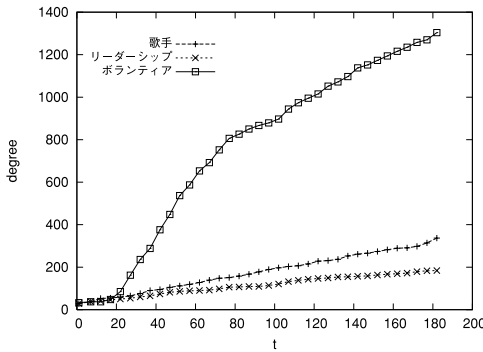


図5 度数の変化の比較

Fig. 5 Comparison of temporal changes in degree.

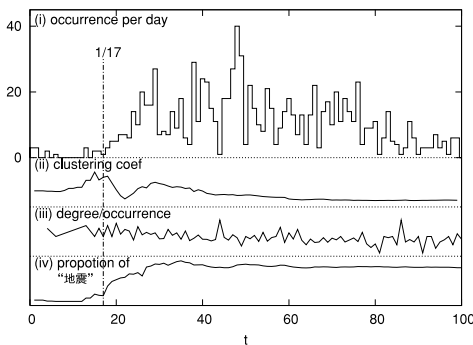


図6 v およびその近傍に関する特徴量の時間変化

Fig. 6 Temporal changes in characteristics of v and its neighborhood.

$\mathcal{K}(t)$ における v のクラスタ係数 .

(iii) 出現頻度あたりの度数の増分

$\mathcal{K}(t)$ における v の度数を k_t とすると, t 日目の度数の増分は $k_t - k_{t-1}$. これを (i) の値で割ったもの .

(iv) “地震” の近傍との重なり具合

$|\Gamma(\text{“地震”}) \cap \Gamma(v)| / |\Gamma(v)|$ の値 .

図5では度数の急激な増加が認められたが, 図6でも v の出現頻度が増加している様子が改めて確認できる. しかし, その値がピークに達するのは震災発生より1カ月ほど経過した後 ($t = 48$) である. 出現頻度あたりの度数の増分 (iii) は, v の出現のうち, 他の語との新たな関係を生成するもの — いい換えれば, 新しい文脈の下で用いられているもの — がどの程度あるのかを示す数量であるが, 全体的に減少傾向にある. また, この出来事と関連性の高い語である “地震” の近傍との重なり具合 (iv) は $t = 35$ でピークに達し, その後, ほぼその値に変化がない. これらの事実を総合すると, v の出現頻度がピークに達している $t = 48$ 付近では, 震災に由来する文脈 (に対応するネットワーク構造) が $\Gamma(v)$ においてすでに形成され

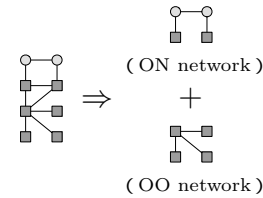


図7 ON ネットワークと OO ネットワークへの分解
Fig. 7 Decomposition of a network into ON and OO networks.

ており, その文脈のもとで v が数多く使われている, と推測できる .

震災に由来する変化の捕捉という観点からすると, 発生直後に変化が認められるクラスタ係数の振舞いが興味深い. 震災後, いったん減少して $t = 22$ に極小値に達するが, その後また増加に転じ, $t = 29$ にピークに達している. t の増加にともない, $\mathcal{K}(t)$ のノード数, リンク数ともに単調に増加するため, ここでのクラスタ係数の減少と増加はそれぞれノード (語) の増加とリンク (語の相互関係) の増加を意味する. よって, v のクラスタ係数の変動は, まず語彙が増加し, その後それらの間に共起関係が発生したことを示していると考えられる .

5.2 ON ネットワーク

$\Gamma(v)$ のクラスタ性の変化をより詳しく調べるため, ここで ON ネットワークという概念を導入する. 共起ネットワークあるいはその部分ネットワーク X に対し, X の ON ネットワーク (Old-New subnetwork) Y を次のように定義する. まず, X のノードのうち, 新しい語彙に属するか, あるいは古い語彙でも新しい語彙とつながりのあるものを Y のノードとする. そして, X のリンクのうち, 新しい語彙に属する語どうし, あるいは新しい語彙と古い語彙をつなげているものを Y のリンクとする. また, X から ON ネットワークのノードを除いて得られるネットワークを OO ネットワーク (Old-Old subnetwork), ON ネットワークの中から, 新しい語彙のみを抽出して得られるネットワークを NN ネットワーク (New-New subnetwork) と定義する. 図7は, ネットワークを ON ネットワークと OO ネットワークに分解した例を示したものである. 分かりやすいように, 古い語彙, 新しい語彙に属するノードを区別し, それぞれ異なった形状 (と) で表示してある. ON ネットワークでは, 古い語彙に属する語どうしのリンクは除かれていることに注意されたい.

図8は, $\Gamma(v)$ を ON ネットワークと OO ネットワークに分解し, 時間経過にともなうノード数とクラ

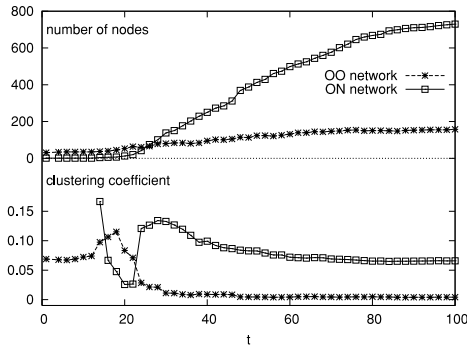


図 8 $\Gamma(v)$ の OO, ON ネットワークの時間変化

Fig. 8 Temporal changes of OO and ON networks of $\Gamma(v)$.

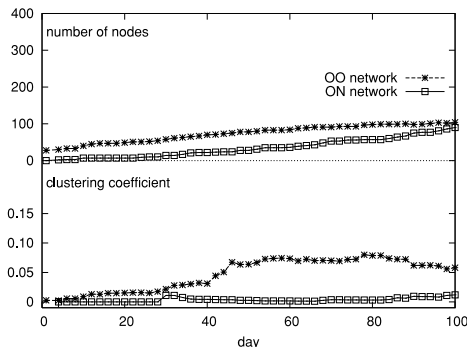


図 9 $\Gamma(\text{“歌手”})$ の OO, ON ネットワークの時間変化

Fig. 9 Temporal changes of OO and ON networks of $\Gamma(\text{“歌手”})$.

スタ係数の変化をそれぞれについて調べたものである。

明らかに、ON ネットワークと OO ネットワークは異なった振舞いを示している。震災前は、 $\Gamma(v)$ において新しい語彙は少数であり、OO ネットワークが優位（ノード数、クラスタ係数ともに他を上回る状態）にある。しかし、震災後およそ 10 日を経て、ON ネットワークは OO ネットワークより優位なネットワークに変化している。この結果は、震災にともなうネットワークの構造的変化を調べるためには、ON ネットワークに注目すべきであることを示唆している。

なお、図 5 で v の比較対象とした“歌手”についても同様に 2 つのネットワークの振舞いを比較してみると、つねに OO ネットワークが優位にあることが分かる（図 9）。

5.3 ON ネットワークの時間変化

すべての語 x に対して、その近傍 $\Gamma(x)$ には、 x がそれ以外のすべてのノードとつながりを持つという共通の構造が存在する。個々の近傍固有の構造を調べるため、本節以降では、この共通構造を除去したネットワークである $\Gamma_*(x)$ (x の近傍から x を除いたもの) に注目し、その ON ネットワークあるいは NN ネット

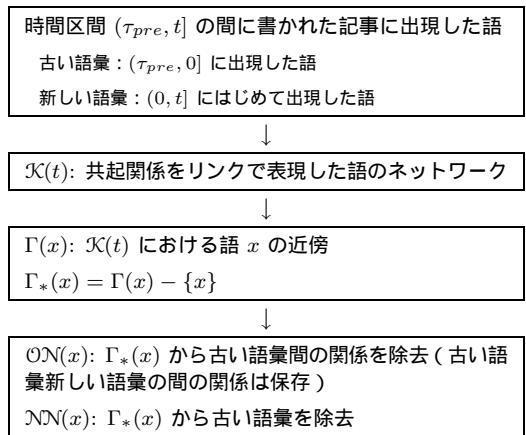


図 10 $\mathcal{ON}(x)$ の生成手順

Fig. 10 Procedure for generating $\mathcal{ON}(x)$.

ワークをもっぱら議論の対象とする。記述を簡潔にするため、これらを「 x の ON (あるいは NN) ネットワーク」と呼び、 $\mathcal{ON}(x)$ (あるいは $\mathcal{NN}(x)$) と表すことにする。ここで、理解の助けとするため、記事データから $\mathcal{ON}(x)$ を作り出す手順を図 10 に整理しておく。なお、定義から、 $\mathcal{ON}(x)$ は、その基となる共起ネットワークの作成時刻 t に依存する。よって、より厳密には $\mathcal{ON}(x, t)$ と時刻 t もあわせて表記すべきであるが、省略しても混乱が生じない場合には $\mathcal{ON}(x)$ と書くことにする。

さて、前節の結果をふまえて、 $\mathcal{ON}(v)$ と $\mathcal{NN}(v)$ の成長を追った（図 11）。 v のクラスタ係数は震災後 $t = 22$ まで減少傾向にあったが（図 6）、この ON ネットワークにおいてもノードの増加が支配的である。リンクの生成数は少なく、単純な構造（鎖状）だけが認められる。その後、 v のクラスタ係数が増加に転じる $t = 24$ では鎖状よりも複雑な構造が形成されはじめ、ON ネットワーク自体のクラスタ係数 も正の値をとる。

新しい語彙に属する語どうしの関係形成、すなわち NN ネットワークの変化は、ON ネットワークに比べて緩慢である。少なくとも $t \leq 29$ の範囲では、ON ネットワークで生成されるリンクのほとんどが新しい語彙内部よりも古い語彙と新しい語彙を結び付けている。それでも、 v のクラスタ係数が極大値に達する $t = 29$ では、NN ネットワークにおいても、鎖状よりも複雑な構造が出現している。ON ネットワークには

本論文では、全ノードのクラスタ係数の平均値をネットワークのクラスタ係数とする。標準的な定義によれば、次数が 2 未満のノードのクラスタ係数は不定となるが、本論文では 0 として計算する。

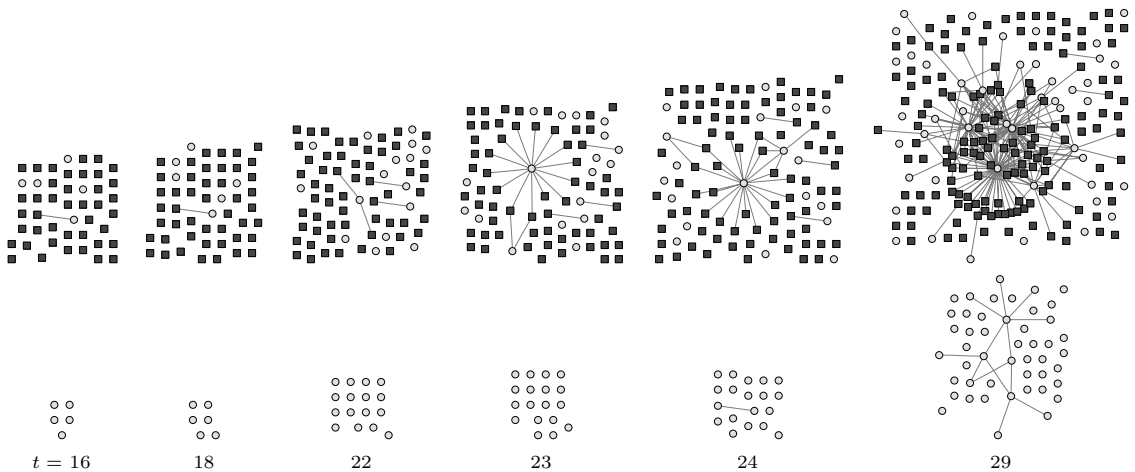


図 11 $\Theta N(v)$ と $NN(v)$ の成長 (上段: ON ネットワーク, 下段: NN ネットワーク)
 Fig. 11 Growing process of $\Theta N(v)$ and $NN(v)$.

古い語彙に属する語どうしを結び付けるリンクが存在しないことに注意すると、実は、新しい語彙内部のリンクは ON ネットワークのクラスタ性を向上させるために不可欠であることが分かる。

5.4 ネットワークの成長とクラスタ性

5.4.1 ノード数とクラスタ性の関係

震災という大きな出来事が発生した後、 $\Theta N(v)$ のクラスタ性が向上する様子を見てきたが、果たしてこの変化は出来事の発生時に特有のものなのだろうか。出来事の発生は短期間のうちに多くの関連語の出現をもたらし (v の) 近傍を成長させる。一方、出来事とは関わりのない語についても、その近傍は、(緩やかではあるが) 時間の経過とともに成長していく。よって、もし、単純に、近傍の成長がクラスタ性の向上につながるのであれば、ノード数の増加を待つことにより、その ON ネットワークのクラスタ性を向上させることができるはずである。

表 2 は v を含む様々な語 x について、 $\Theta N(x, t)$ の特徴量を計算したものである。 t, n, l, C はそれぞれの共起ネットワークの作成時刻、 $\Theta N(x, t)$ のノード数、リンク数、そしてクラスタ係数である。作成時刻 t は、 n が互いにほぼ等しくなるように語ごとに選んだ。ネットワークの大きさ n がほぼ等しいにもかかわらず、“歌手”と“リーダーシップ”にはクラスタ性が認められない (クラスタ係数が 0 である)。つまり、この例では成長を待ってもクラスタ性が向上しないのである。一方、阪神淡路大震災に関連性のある“ライフライン”、1995 年 3 月 20 日 ($t = 79$) に起きた地下鉄サリン事件に係る“サリン”、“不審物”の ON ネットワークも調べた結果、いずれも正のクラ

表 2 $\Theta N(x, t)$ のクラスタ係数の比較

Table 2 Comparison of $\Theta N(x, t)$ in clustering coefficient.

x	t	n	l	C
歌手	151	55	45	0.0
リーダーシップ	187	57	79	0.0
ボランティア	27	55	60	0.078
ライフライン	38	56	88	0.196
サリン	82	61	72	0.024
不審物	188	50	64	0.022

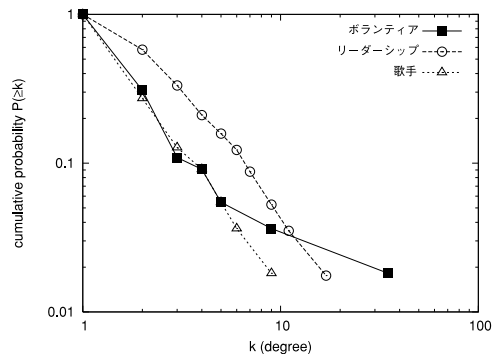


図 12 $\Theta N(x)$ の度数分布の比較

Fig. 12 Comparison of degree distributions of $\Theta N(x)$.

スタ係数が得られた。

5.4.2 リンク数とクラスタ性の関係

一般的に、リンク数の増大はクラスタ性の向上に寄与すると考えられる。しかしながら、表 2 に示した結果からは、リンク数の多少とクラスタ性の有無に関連性を見出すことができない。

そこで、ノードごとのリンク数の分布、すなわち度数分布を比較した。図 12 は、表 2 に示した 6 つのネットワークから ΘN (“リーダーシップ”, 187),

ON(“歌手”, 151), ON(“ボランティア”, 27) を選び、次数の累積分布を両対数スケールで示したものである。グラフから、これらのネットワークには、次数の低いノードが大半を占め、次数の高いノードが少数存在しているという共通の性質があることが見て取れる。

しかし、両者の分布形に違いがある。“リーダーシップ”や“歌手”では、全体的に次数の存在確率がベキ的に減衰している（すなわち、 $P(\leq k) \propto k^{-\gamma}$ ）。この分布（ベキ分布）を持つネットワークはスケールフリーネットワーク¹⁾と呼ばれており、様々なネットワークがこの範疇に属していることが分かっている。一方、“ボランティア”では、低次数の範囲($k \leq 5$)のベキ的減衰から見積もられるよりも高い次数ノードが存在している。ベキ分布に従うのが共起ネットワークの「通常の振舞い」であるとするならば、“ボランティア”では高次数への逸脱が起きている、ととらえることができるだろう。この結果から、出来事の発生はノードの増加とともに、局所的なリンクの増加、いい換えれば、特定の語との「過度」に集中的な共起関係の増加をもたらしており、それがクラスタ構造を出現させている、と推測できる。

5.5 τ_{pre} への依存性

共起ネットワーク $\mathcal{K}(t)$ は、時間区間 $(\tau_{pre}, 0]$ に出現した古い語彙と、 $(0, t]$ に出現した新しい語彙から構成されている。よって、新しい語彙を主な要素とする ON ネットワークの構造は t だけでなく τ_{pre} の選び方に依存する。いま、 t を固定して τ_{pre} を増加（絶対値を減少）させると、古い語彙が減り、その減少分だけ新しい語彙が増加する。その結果、新しい語彙を主たる要素とする ON ネットワークのサイズが増大し、そこにクラスタ性が出現する可能性も高くなると思われる。

この予想を確認するため、 $\tau_{pre} = -14$ という条件のもとでは ON ネットワークにクラスタ性が認められなかった“歌手”と“リーダーシップ”について、 τ_{pre} への依存性を調べた。その結果を図 13 に示す。このグラフは、各 τ_{pre} ごとに共起ネットワーク $\mathcal{K}(180)$ を構成し、そこからそれぞれの語 x について ON(x) を抜き出し、そのノード数 (n) とクラスタ係数 (C) を調べ、プロットしたものである。なお、ここでは、個々の正確な値を知ることよりも変動傾向の把握を目的としているため、グラフが見やすくなるよう各々適宜スケールを変えてある（図には、「 $\times 10$ 」のように倍

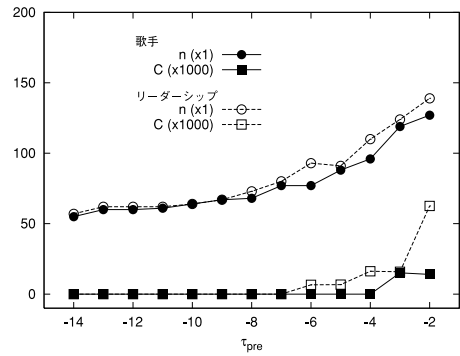


図 13 ネットワーク構造の τ_{pre} への依存性

Fig. 13 Dependency of network structures on τ_{pre} .

表 3 1991 年の記事を土台とした場合のクラスタ係数

Table 3 Clustering coefficient of networks with base terms appearing in articles in 1991.

x	t	n	l	C
歌手	151	77	71	0.0
リーダーシップ	187	89	165	0.0
ボランティア	27	50	56	0.040
ライフライン	38	51	70	0.154
サリン	82	75	92	0.022
不審物	188	58	100	0.108

率を記した)。

いずれの語においても、予想どおり、 τ_{pre} の増加に伴わないクラスタ性が出現している。しかし、その振舞いは、 n の変動に見られるような単調な増加ではない。 C は、 τ_{pre} が増加しているにもかかわらず、 $\tau_{pre} < -6$ まで変化がなく 0 の状態が続いている。つまり、ある程度範囲内（この例では、 $-14 \leq \tau_{pre} \leq -7$ ）では、これらの語の ON ネットワークの構造は τ_{pre} に非依存で安定していることが分かる。

なお、古い語彙と新しい語彙の間に時間的な連続性を要求しないのであれば、共起ネットワーク $\mathcal{K}(t)$ の「土台」となる $\mathcal{K}(0)$ をさらに別な方法で構成することも考えられる。一例をあげると、 $t = 0$ に対応する基準日 T_0 とは無関係にある日を選び、そこからある日数の間に書かれた記事により共起ネットワークを構成し、 $\mathcal{K}(0)$ の代わりとするのである。表 3 は、1991 年 12 月 18 から同年 12 月 31 日までの 2 週間に書かれた記事をもとに $\mathcal{K}(0)$ の代わりとなる共起ネットワークを構成し、表 2 と同じ特徴量を測った結果である。ここでも、表 2 同様、大きな出来事に関係する語のみにクラスタ性が認められている。

6. 語のバーストと共起ネットワークにおけるクラスタ構造の形成

一般に、震災のような出来事の発生は、新聞記事の

神経細胞、代謝、食物連鎖、論文の共著作、企業間の株の保有関係などを表すネットワークを例としてあげることができる。

文書ストリームにおいて、それに関係する語のバーストをもたらす。一方、前章では、共起ネットワーク（の一部である ON ネットワーク）において、出来事に由来してクラスタ構造が現れることをいくつかの例で確認した。本章では、出来事とクラスタ構造の関連性を詳しく調べるため、バーストを実社会の出来事に対応するものとしてとらえ、クラスタ構造との関係を定量的に解析する。

6.1 バーストの検出

文書ストリームにおけるバーストの検出については、今までに複数の方法が提案されている^{5),6),8),16)}。本論文では、これらの中から Kleinberg の手法⁸⁾を採用した。その理由は、計算のコスト（処理時間や計算リソース）が低く、かつ、本論文が対象とするデータに対して妥当な結果を導き出すことができたからである。

Kleinberg の手法は次のような考えに基づいている。いま、時刻 t ごとに d_t 個の記事が書かれ、そのうち r_t 個に語 x が出現している状況を考える。直感的には、 r_t/d_t が高い値を示すとき、 x （の出現状況）は時刻 t においてバースト状態にあると考えられるのだが、ゆらぎの結果としてこの値が一時的に高くなることもある。そこで、文書ストリームを通常状態とバースト状態の2つの状態からなるオートマトンでモデル化し、観測値 d_t, r_t と推定状態との適合性を計算することにより各時刻における状態を決定する。さらに Kleinberg は、このモデルに基づき、与えられた時間区間における x のバーストの度合いを表す数量を定義した。本論文では、これをバースト度と呼び、 $\beta(x)$ で表すことにする。

この手法を 1995 年 1 月 1 日 ($t = 1$) からの 180 日間に書かれた新聞記事の文書ストリームに適用してみたところ、阪神淡路大震災 ($t = 17$) に関係のある“ボランティア”では時間区間 [25, 75] と [121, 124] に、地下鉄サリン事件 ($t = 79$) に関係のある“サリン”では時間区間 [79, 164] にバースト状態にあるという結果が得られた。また、この期間に最もバースト度の高い語は“阪神大震災”であった。

6.2 クラスタ性のバースト性との関係

6.2.1 $ON(x)$ のクラスタ性と x のバースト性

Kleinberg の手法を用いて、時間区間 [1, 180] に記事に現れた語 x のバースト度 $\beta(x)$ を計算し、ネットワーク $ON(x, 180)$ のクラスタ性との関係を調べた。表 4 は、クラスタ性とバースト性 ($\beta(x) > 0$ である

表 4 クラスタ性とバースト性の関係を示す分割表
Table 4 Crosstable relationship between clustering property and burstness.

	$\beta(x) > 0$	$\beta(x) = 0$	
$C(ON(x)) > 0$	3253	3749	7002
$C(ON(x)) = 0$	3775	534351	538126
	7028	538100	545128

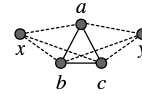


図 14 クラスタ構造の共有

Fig. 14 Neighborhoods sharing a cluster structure.

ということ)という2つの性質の関係を示す分割表である。ここで、 $C(X)$ はネットワーク X のクラスタ係数を表す。なお、 $ON(x)$ のもととなる共起ネットワーク $\mathcal{X}(180)$ には時間区間 $(\tau_{pre}, 0]$ に現れた語も含まれるが、ここでは、時間区間 [1, 180] に出現した語のみを比較対象としている。

χ^2 検定によれば、クラスタ性とバースト性には有意な関連性があることが分かる。しかし、この表からは、(因果関係のような)明確な呼応を読み取ることはできない。たとえば、クラスタ性を持つ語のうちバースト性があるものの割合は $0.466 (= 3253/7002)$ であり、過半数はバースト性と関係がないという結論が導かれる。

6.2.2 比較方法の再検討

ここで、「 $ON(x)$ のクラスタ性と x 自体のバースト性との関連付ける」というアプローチを見直してみたい。たとえば、いま、図 14 に示したように、ノード x の近傍がノード a, b, c で構成されるクラスタ構造 Δ を含んでおり、さらにこの Δ は、 x と異なるノード y の近傍にも含まれている状況を考える（クラスタ構造と x, y をつないでいるリンクは破線で示してある）。このとき、仮に x 自身にはバースト性がなくても、 y にバースト性があるならば Δ はバーストに関連していると見なす、というのも自然な考えだろう。

そこで、この考え方に基づいてクラスタ性とバースト性の関係を再検証した。その具体的な内容は以下のとおりである。すなわち、6.2.1 項では、

$$\mathcal{C} = \{x | C(ON(x)) > 0\}$$

$$\mathcal{B} = \{x | \beta(x) > 0\}$$

という二者の関係を調べたが、これに対し、ここでは \mathcal{C} と、

$$\tilde{\mathcal{B}} = \{x | \exists y \in \mathcal{B}, C(ON(x) \cap ON(y)) > 0\}$$

度合いに応じた複数のバースト状態を持つオートマトンとしてモデル化することもできる。

表 5 クラスタ性とバースト性の関係

Table 5 Relationship between clustering property and burstness.

年	$ K $	$ C $	$ B $	$ B $	$ C \cap B / C $
1995	557463	7002	7028	10118	0.906
1994	522960	5194	6664	8526	0.810

表 6 クラスタ性とバースト性の関係 (出現頻度別)

Table 6 Relationship between clustering property and burstness (by frequency).

出現頻度 f	$ K $	$ C $	$ B $	$ C \cap B / C $
$0 \leq f < 100$	540687	3867	6684	0.866
$100 \leq f < 200$	2118	1096	1357	0.936
$200 \leq f < 300$	759	563	594	0.933
$300 \leq f < 400$	441	388	388	0.943
$400 \leq f < 500$	276	256	257	0.969

の関係調べるのである。その結果を表 5 に示す。この表には、1994 年の 1 月 1 日から 180 日間に出現した語について調べた結果もあわせて載せてある。 $|C \cap B|/|C|$ の列が、クラスタ性とともバースト性が認められる割合であり、1994 年、1995 年のいずれも高い値となっている。

さらに、表 6 には 1995 年のデータを語の出現頻度ごとに分けて調べた結果を示す。出現頻度が 100 未満の語が全語彙の中に占める割合は 99.2% であり、ほとんどの語がこの範囲に属していることが分かる。この中でクラスタ性が認められるものは 0.7% であり、これは C 全体の 55.2% にあたる。そして、このうち 86.6% にバースト性との関連が認められる。

6.2.3 バースト度とクラスタ性

前項では、ほとんどのクラスタ構造がバーストに関連していることを示した。しかし、この逆は必ずしも成り立たない。つまり、Kleinberg の手法で語 x にバースト性があると判断されながらも $\mathcal{O}N(x)$ にはクラスタ構造が見つからない場合が少なからずある。図 15 はその状況を示したもので、語 x を $\beta(x)$ について降順に順位付けし、100 位ごとに $C(\mathcal{O}N(x)) > 0$ となるものの割合を計算し、ヒストグラムで表したものである。

表 4 に示したように、バーストの半数以上がクラスタ構造をとまなわれないことが分かっていたが、このグラフから、バースト性が弱まるにつれ、クラスタ性の出現率が低下していることが分かる。逆のいい方をすれば、バースト性が強い場合にはクラスタ性との関連性も高くなる。たとえば、1~300 位に属する語では、その 9 割以上にクラスタ性が認められている。

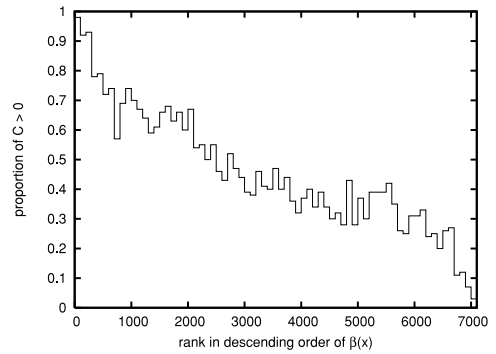


図 15 バースト度とクラスタ性の関係

Fig. 15 Relationship between intensity of burstness and clustering property.

7. 出来事との対応付け

前章での議論から、語のバースト性とクラスタ性の対応関係が明らかになった。一方、阪神淡路大震災の関連語に“ライフライン”や“ボランティア”があるように、一般に、出来事には関連語が複数存在する。よって、極端な場合には、バースト性の高い語がごく限られた出来事の関連語で占められている可能性もある。もしそうであれば、共起ネットワークにおけるクラスタ構造は、ごく限られた出来事に対応していることになる。そこで、本章では、出来事とクラスタ構造の対応を確認する。具体的には、ここまで例としてあげた阪神淡路大震災や地下鉄サリン事件だけでなく、社会的に認知されているそのほかの出来事に対しても、ある関連語が存在してバースト性とクラスタ性が認められることを示す。

表 7 は、1995 年前半に発生した主な出来事と、それを表す語 x 、同期間 (1 月 1 日から 180 日間) に書かれた記事から作り出された共起ネットワーク $K(180)$ における x のクラスタ性 (正確には、 $\mathcal{O}N(x, 180)$ のクラスタ性) の有無を示したものである。出来事は平凡社の 1996 年版百科年鑑⁷⁾において、「12 大インデックス」としてあげられているものから、当年前半に起きたものを選び出したものである。関連語の抽出方法は以下のとおりである。まず、年鑑における出来事の記述中にある名詞およびその複合語を取り出し候補とする。次に、その中からバースト性のある語を選び出す。その結果、出来事あたり 1 語のみに絞られた場合にはそれを関連語とする。一方、複数の語が得られた場合には、著者の主観的判断で出来事を最も端的に表しているものを関連語として採用した。この方法により、いずれの出来事に対してもバースト性を有する妥当な関連語を得ることができた。

表 7 1995 年前半の主な出来事を表す語とその共起ネットワークにおけるクラスタ性
Table 7 Terms representing major incidents in the first half of 1995 and their clustering properties in the co-occurrence network.

出来事	発生時期	関連語 x	$\mathcal{ON}(x)$ のクラスタ性
阪神淡路大震災	1/17	阪神大震災	
オウム関連事件	3/20~	地下鉄サリン事件	
フランスによる核実験再開	6/13~	核実験再開表明	x
戦後五十年国会決議	6/9	戦後五十年国会決議	x
野茂, 大リーグで活躍	5/2~	野茂投手	
金融機関の破綻	(通年)	不良債権処理	
無党派の青島・横山知事当選	4/9	統一地方選	
水俣病問題に最終解決案	4/28~	水俣問題	

さて、表にあげた主要な出来事 8 件のうち 6 件では、出来事のジャンルにかかわらず、クラスタ性との関連が示されている。残りの 2 件は、いずれも出来事の発生が前半期の終わりに近い(6 月中)という共通性がある。この場合、これらは前半期に限ると規模の小さい出来事としてとらえられてしまう可能性がある。具体的には、これらの出来事に関連する記事のうち、 $K(180)$ の構成に寄与するものの数が少なく、その結果、クラスタ構造の出現に至らないということが起こりうる。このような「矮小化」の可能性がある場合を除き、表 7 にあげた出来事はいずれも、共起ネットワークにおけるクラスタ構造と対応付けられている。

8. 本手法の性質と適用限界

本手法の性質の 1 つに、情報提供方法(スタイル)への依存性がある。

極端な例ではあるが、一般には大きな出来事として扱われている阪神淡路大震災も、それが記事になるべき出来事として扱われるか否かは書き手の判断による。そして、記事にならなければ、当然共起ネットワークにはクラスタ構造は生じない。つまり、クラスタ構造が生み出されるか否かは情報提供スタイルに依存しているのである。本論文では、高い公共性が求められている新聞というメディアからの情報を用いることで客観性の確保を狙った。そして、実際、社会的に認知されている出来事を語のバーストとしてとらえることができ、さらにそれをクラスタ構造に対応付けることもできた。

スタイルに一貫性があれば、記事作成の時期に依存せずに、本手法が適用できると考えられる。実際、6.2.2 項で示したように、1994 年および 1995 年それぞれに作成された記事を対象に解析した結果、双方で

バースト性とクラスタ性の間に対応関係が認められている。

スタイルが異なる場合、たとえば、異なる新聞社、あるいは異なる言語で書かれた記事を用いた場合に出来事のとりえ方に違いが生じるのか否かという問題は興味深い。少なくとも前者(同種のメディアで同一言語の場合)では、本手法の適用が可能であると期待される。一方、Web のような多数の情報源がそれぞれ自律的に情報を発信しているメディアでは状況が大きく異なる。それゆえ、本手法の適用可能性は現時点では不明である。もし、情報提供スタイルの違いが本手法に影響を及ぼすのであれば、どのようなかたちで作用するのかという問題もまた興味深いものであり、今後取り組むべき研究課題である。

9. ま と め

共起ネットワークにおける語 x の近傍は、文書において x に近接して出現する語の集まりであり、 x が用いられている文脈を表していると考えられる。 x が用いられる文脈に何らかの変化が発生したとき、その近傍にも構造的変化が生じる。本論文では、まず、文脈の変化を発生させる原因として、阪神淡路大震災という出来事を題材にとりあげ、新聞記事から生成される共起ネットワークの変化の様子、特に、“ボランティア”という語の近傍の構造的変化を調べた。その結果、相対的に新しい語彙を主な構成要素とする部分ネットワーク $\mathcal{ON}(x)$ (語の近傍の ON ネットワーク)の性質をとらえることが重要であることが分かった。そして、その部分ネットワークのクラスタ性に出来事の影響を見出せることを複数の実例を用いて示した。

さらに、出来事とクラスタ性との関連性を詳しく調べるため、 $\mathcal{ON}(x)$ のクラスタ性と、出来事に付随して発生する語の高頻度な出現(バースト)の関係を調べた。その結果、共起ネットワークの(局所的な)成長を $\mathcal{ON}(x)$ の構造上の特徴を基準として 2 つのタイプに分類すると、それぞれ以下のような特徴を持つこ

たとえば、フランスによる核実験再開については、その再開表明は 6 月であったが、国際的な反対にもかかわらず同年 9 月以降に数回にわたって核実験を強行したという事実が大きな問題としてとりあげられている。

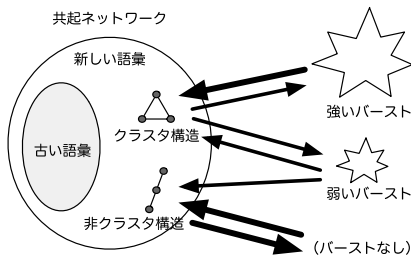


図 16 クラスタ構造とバーストの関係を示す模式図

Fig. 16 Schematic representation of the relationship between clustering structures and bursts.

とが分かった。

(i) $\mathcal{ON}(x)$ にクラスタ構造が出現しないもの

このタイプでは、語のバーストとの関連性は非常に低い。つまり、このタイプは、文書ストリームにおいて一定の割合で文書が生産される続けることにより生じる緩やかな成長である。

(ii) $\mathcal{ON}(x)$ にクラスタ構造が出現するもの

このクラスタ構造は、ほとんどの場合、バースト性のある語の近く（近傍）に出現している。このことから、このタイプにはバーストの関与が示唆される。実際、強いバーストが発生したときには、当該語の近傍の成長は高い確率でこのタイプとなる。

この結果をまとめると図 16 のようになる。バーストを出来事に対応するものととらえれば、 $\mathcal{ON}(x)$ に出現しているクラスタ構造は何らかの出来事の形跡と考えられる。特に、大きな出来事のほとんどは共起ネットワークにクラスタ構造として刻み込まれている。

バーストが発生したとき、共起ネットワークも（局的に）急激に成長する。この成長はタイプ (i) の成長速度を速めたものとは本質的に異なり、その違いがクラスタ構造を生み出すと考えられる。そのダイナミクスの詳細については現時点ではまだ分かっていないが、バーストにともないリンク（共起関係）の生成が特定ノードに「過度」に集中していることがいくつかの例で確認されており（5.4.2 項）、これがクラスタ構造の出現につながるのではないかと推測される。

参 考 文 献

- 1) Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol.286, pp.509–512 (1999).
- 2) Dorogovtsev, S.N. and Mendes, J.F.F.: Language as an evolving word web, *Proc. Royal Society of London Series B*, Vol.268, No.1485, pp.2603–2606 (2001).
- 3) Ferret, O.: Discovering word senses from a

network of lexical cooccurrences, *Proc. 20th international conference on Computational Linguistics*, pp.1326–1332 (2004).

- 4) Ferrer i Cancho, R. and Solé, R.V.: The small world of human language, *Proc. Royal Society of London, Series B*, Vol.268, No.1482, pp.2261–2266 (2001).
- 5) Fujiki, T., Nanno, T., Suzuki, Y. and Okumura, M.: Identification of Bursts in a Document Stream, *Proc. 1st International Workshop on Knowledge Discovery in Data Streams*, pp.55–64 (2004).
- 6) Fung, G.P.C., Yu, J.X., Yu, P.S. and Lu, H.: Parameter Free Bursty Events Detection in Text Streams, *Proc. 31st International Conference on Very Large Data Bases*, pp.181–192 (2005).
- 7) 平凡社百科年鑑 '96 (1996).
- 8) Kleinberg, J.: Bursty and Hierarchical Structure in Streams, *Proc. 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.91–101 (2002).
- 9) 毎日新聞社：CD-毎日新聞 91～95 版。
- 10) 増田直紀，今野紀雄：複雑ネットワークの科学，産業図書 (2005)。
- 11) Matsuo, Y., Ohsawa, Y. and Ishizuka, M.: Document as a Small World, *Proc. JSAI 2001 International Workshop (LNAI 2253)*, pp.444–448 (2001).
- 12) Newman, M., Barabási, A.-L. and Watts, D.J. (Eds.): *The Structure and Dynamics of Networks*, Princeton University Press (2006).
- 13) 丹羽芳樹：動的な共起解析を用いた対話的文書検索支援，情報処理学会研究会報告，96-NL-115, pp.99–106 (1996).
- 14) Ohsawa, Y., Benson, N.E. and Yachida, M.: KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor, *Proc. IEEE Forum on Research and Technology Advances in Digital Libraries*, pp.12–18 (1998).
- 15) Solé, R.V., Murta, B.C., Valverde, C. and Steels, L.: Language Networks: Their Structure, Function and Evolution, Santa Fe Institute Working Paper, 05-12-042 (2005).
- 16) Swan, R.C and Allan, J.: Extracting Significant Time Varying Features from Text, *Proc. International Conference on Information and Knowledge Management*, pp.38–45 (1999).
- 17) 高田寛喜，斉藤和巳，木村昌弘：共起ネットワークの成長分析，人工知能学会研究会資料，SIG-KBS-A504-21, pp.117–122 (2006)。
- 18) Takano, A., Niwa, Y., Nishioka, S., Iwayama, M., Hisamitsu, T., Imaichi, O. and Sakurai,

H.: Associative Information Access Using Dual-NAVI, *Kyoto International Conference on Digital Libraries2000*, pp.285-289 (2000).

- 19) Tanaka-Ishii, K. and Iwasaki, H.: Clustering Co-occurrence Graph based on Transitivity, *Proc. 5th Workshop on Very Large Corpora*, pp.91-100 (1997).
- 20) Veling, A. and van der Weerd, P.: Conceptual Grouping in Word Co-Occurrence Networks, *Proc. 16th International Joint Conference on Artificial Intelligence*, pp.694-701 (1999).
- 21) Watts, D.J. and Strogatz, S.H.: Collective dynamics of small-world networks, *Nature*, Vol.393, pp.440-442 (1998).
- 22) Whittaker, S., Isaacs, E. and O'Day, V.: Widening the Net: Workshop Report on the Theory and Practice of Physical and Network Communities, *SIGCHI Bulletin*, Vol.29, No.3, pp.27-30 (1997).

(平成 19 年 3 月 20 日受付)

(平成 19 年 7 月 3 日採録)

(担当編集委員 酒井 哲也)



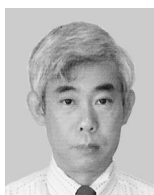
佐藤 進也 (正会員)

1988 年東北大学大学院理学研究科数学専攻修士課程修了。同年日本電信電話株式会社入社。以来、協同作業支援、Web 情報検索・マイニング等の研究に従事。複雑系の情報処理とその応用に興味を持つ。現在、NTT 未来ねっと研究所主任研究員。博士(情報理工学)。訳書『スモールワールド』(ダンカン・ワッツ著、東京電機大学出版局、共訳)。ACM、ISOC、電子情報通信学会各会員。



福田 健介

1999 年慶應義塾大学大学院理工学研究科計算機科学専攻後期博士課程修了(博士(工学))。同年日本電信電話株式会社入社。以来、未来ねっと研究所に所属。この間 2002 年ボストン大学訪問研究員。2006 年より国立情報学研究所アーキテクチャ科学研究系准教授。学術情報ネットワーク、インターネットおよびネットワーク科学に関する研究に従事。



菅原 俊治 (正会員)

1982 年早稲田大学大学院理工学研究科数学専攻修士課程修了。同年日本電信電話株式会社入社。以来、知識表現、学習、マルチエージェントシステム、インターネット等の研究に従事。1992~1993 年マサチューセッツ大学アムハースト校客員研究員。2007 年より早稲田大学基幹理工学部情報理工学科教授。博士(工学)。日本ソフトウェア科学会、電子情報通信学会、人工知能学会、ISOC、IEEE、ACM、AAAI 各会員。



栗原 聡 (正会員)

1992 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了。同年日本電信電話株式会社入社。2004 年から大阪大学大学院情報科学研究科/大阪大学産業科学研究所知能システム科学研究部門准教授。分散人工知能、インターネット、ネットワーク科学等の研究に従事。博士(工学)。日本ソフトウェア科学会、人工知能学会各会員。