Vol.2012-MPS-87 No.35
2012/3/2

# クリーク全列挙に基づく構造変化検出アルゴリズム

エラウィンディ サラ[†1]　原口　誠[†1]
大久保　好章[†1]　富田悦次[†2]

本稿では，高次数の独立集合が孤立度が高いクリークに変化する場合を，グラフ構造変化の特徴的なパターンと考え，これらを枚挙する手法をクリーク全列挙手法の拡張として与える．

# Detecting Structural Change of Graph Based on Constrained Maximal Clique Search

Sarah El-Lawindy,[†1] Makoto Haraguchi,[†1]
Yoshiaki Okubo[†1] and Etsuji Tomita[†2]

This report discusses a problem of structural change detection given two graphs before and after some change, targeting vertex sets that are divergent independent sets before the change and maximal cliques with smaller number of outgoing edges after the change. A search algorithm for them is presented based on a constrained maximal clique enumeration algorithm.

## 1. はじめに ー ターゲットとするもの

目まぐるしく変化する情報の世界においては，変化および変化の兆しを検出するのは重要なタスクであると考えられる．顕著な変化を検出するシステムとしては，emerging pattern[7]，contrast set[8]，バースト解析[6] など様々なアプローチがあるが，顕著ではない変化も検出できる研究は未だ十分ではないと思われる．偶発的変化も変化としては多数存在するからである．したがって，顕著でない変化のうち，偶発的なものを排除し，ある程度の必然性を持つものに絞りこむことが重要である．この立場に立つ研究として，トランザクションデータベースに対する相関変化の研究がある[4),5)]．一方，トランザクションデータのみならず，グラフの形で与えられるものも多数存在し，イベント発生の前後，トピック，時間，カテゴリー等の文脈変化の前後における２つのグラフにおいて，潜在的だが決して偶発的ではない変化の候補を検出するタスクは意味があると考える．

無向グラフに対する頂点結合の必然性（もしくは逆に偶発性）の度合いを図る尺度として モデュラリティ[3] が提案され，グラフ分割・クラスタリング等において活発に使われている．本稿では モデュラリティ尺度を直接用いないが，その考えかたに従い，モデュラリティが高くなる一つの構造的条件がクリークであると考え，構造変化を検出するための一方式を提案する．

モデュラリティでは現実のグラフとランダムな場合との差異を観測する．具体的には，ランダムグラフにおいて頂点 $i,j$ が辺で結合される期待値 $P_{ij}$ を算出する．ただし，辺の生成は一様確率に基づく独立試行列ではなく，所与のグラフにおける次数を（期待値の意味で）保存する確率の中で，最もランダムなものを用いたときの独立試行列による辺生成を行う．導出される $i,j$ 間の期待値 $P_{ij}$ は $\frac{deg(i)deg(j)}{2m}$ である．ここに，$deg(i)$ は頂点 $i$ の次数，$m$ は所与のグラフにおける辺の総数である．隣接頂点 $i,j$ 間の モデュラリティは $1-P_{ij}$，非隣接の場合は，$-P_{ij}$ で定め，一般には正および負の値をとる．低次数の頂点同士が結合されているほど，モデュラリティは正の高い値となり，高次数の頂点同士が非結合の場合ほど，モデュラリティは負の小さな値（絶対値大）となる．これは，低次数のものがその結合力の弱さにも関わらず隣接関係にある場合は必然性があり（正），また，高い結合力を持つ高次数のものが非結合の場合は必然性が低い（負）と解釈できる．

こうした モデュラリティに基づき，変化の前後における偶発的でない変化を計測する一つの方法として，変化前の モデュラリティと 変化後の モデュラリティの差分をとり，差異を大にする頂点組合せを求める方法[9] も提案されている．本稿では，変化の前後における モデュラリティの差異が高くなる構造上の典型的な特徴に着目し，構造変化検出問題としてとらえる．具体的には，図 1 に示されるように，変化前においては高次数の頂点が非結合であり，独立集合であることを要請する．定義から モデュラリティは負の値をとり，次数が高いほどより小な負となる．一方，変化後においては，低次数の頂点から構成されるクリークに変化し，外部への接続が少数なことを要請する．小規模なクリークで外部に出る辺が

†1 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University
†2 電気通信大学先進アルゴリズム研究ステーション
Advanced Algorithm Research Laboratory, The University of Electro-Communications

少数なことから，頂点の次数はグラフが持つ辺の総数よりもはるかに小な値となり，モデュラリティは正の比較的に高い値をとりやすい．ここで，変化後のクリークの頂点数が大な場合は，モデュラリティは相対的に低くなるが，大規模な密結合部分グラフは，極大クリーク列挙手法やグラフ分割手法によって抽出可能であり，その理由により本稿では（変化後に）比較的に小規模のクリーク検出を想定している．

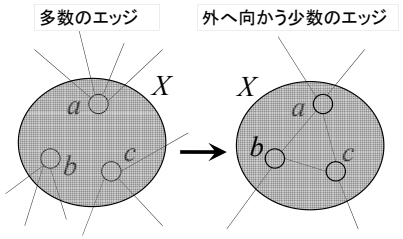


**図 1** 発散的な独立集合から内向きのクリークへ

## 2. 統合グラフとその評価

本節では，統合グラフとその評価について述べる．目標とするパターン，すなわち頂点集合には，変化前の補グラフと変化後のグラフにおいて共にクリークをなすことを要請する．このことを，統合グラフのクリークとして表現する．また，パターンからその外部への接続状況を変化の前後において評価する．部分グラフとしての構造制約は統合グラフの構造制約として，接続状況に関する制約は，変化前後のグラフに依存した評価により与える．特に，評価は，分枝限定法を用いた目的関数値の見積もり法と同様に，追加可能な候補頂点集合も含めた形で行う（暫定評価）．

### 2.1 統合グラフ

本研究で与えられるグラフは，Newman 論文[3] 同様に，重みなしの (単純) 非有向グラフであり，ただし，コントラストをとるための変化前のグラフ $G_1 = G_1(V_1, \Gamma_1)$ と 変化後のグラフ $G_2 = (V_2, \Gamma_2)$ が与えられるとする．ここで $V_j$ は頂点集合，$\Gamma_j \subseteq V_j \times V_j$ はエッジ集合であり，各 $x \in V_j$ の隣接頂点集合を $\Gamma_j(x)$ で表す．

目標となるパターンとは，$G_j$ $(j = 1, 2)$ の共通の頂点集合 $X \subseteq V_1 \cap V_2$ であり，$X$ の任意の頂点対が変化前において非隣接かつ，変化後において隣接関係にあることを要請する．すなわち，$G_1$ において独立集合，$G_2$ においてクリークをなす．つまり，$G_1$ の補グラフのエッジ集合を $\Gamma_1^c = (V_1 \times V_1) \setminus (\Gamma_1 \cup \{(x, x) | x \in V_1\})$ とすると，$X$ は統合グラフ $G = (V = V_1 \cap V_2,\ \Gamma = \Gamma_1^c \cap \Gamma_2)$ におけるクリークである．

節 1 で述べたように，本論文では，モデュラリティが高くなる典型的な部分グラフを優先的に抽出するアルゴリズムの設計を目標とする．「典型性」の定義にはいくつかのバリエーションがあるが，本稿ではクリーク内部から外部に出ていくエッジ数を評価する方式を採用する．その他の方法について，最終節 4 において論じる．

### 2.2 変化後のパターン評価

変化後の頂点集合 $X$ は，相互に結合し（クリーク）かつ $X$ 以外への接続が（次数に照らして）少ないものが，より変化の後で専門性の高い頂点のコミュニティが形成されていると解釈する．クリークをなすことは，統合グラフ $G$ における要請であり，また，外部との接続関係は，変化後のグラフ $G_2$ における性質であることに注意し，下記で定める．

**定義 2.1** $G$ **の極大クリーク** $X$ **の評価**：$X$ 中の頂点から外部に出るエッジ率の最大値

$$E_2(X) = \max_{x \in X} \frac{outgoing_2(x|X)}{deg_2(x)}, \quad \text{where}$$

$G_2$ における $x$ の次数：$deg_2(x) = |\Gamma_2(x)|$,

外部接続エッジ数：$Y \subseteq V_2$ に対し，$outgoing_2(x|Y) = |\Gamma_2(x) - Y|$

クリークは一般に極大クリークの部分グラフとして出現することを考慮し，上記定義は，極大クリークの評価のために用いる．一方，後述するクリーク全列挙アルゴリズムにおいては，暫定的なクリーク $X$ に対する暫定評価 $est\text{-}E_2(X)$ を行う．

**定義 2.2** $G$ **のクリーク** $X$ **の暫定評価**：

$$est\text{-}E_2(X) = \max_{x \in X} \frac{outgoing_2(x|X \cup Cand(X))}{deg_2(x)}, \quad \text{where}$$

$X$ に追加可能な候補頂点集合 $Cand(X) = \{y \in V \mid X \subseteq \Gamma(y)\}$

**事実 2.1** (1) $G$ の極大クリーク $X$ に対し，$E_2(X) = est\text{-}E_2(X)$.
(2) $G$ のクリーク $X_1, X_2$ に対し，$X_1 \subseteq X_2 \Rightarrow est\text{-}E_2(X_1) \leq est\text{-}E_2(X_2)$

$G$ の極大クリーク $X$ に対し，$Cand(X) = \phi$. よって，(1) は明らか．また，$X_1 \subseteq X_2$ なる $G$ のクリーク $X_1$，$X_2$ に対し，$X_1$ に候補を追加してできる $G$ のクリーク列 $X_1 = Y_0, Y_1, \cdots, Y_n = X_2$ を構成できる．各拡張段階 $Y_{k+1} = Y_k \cup \{u_k\}$ （$u_k \in Cand(Y_k)$）に

おいて，クリークと候補頂点集合を併せた $Y_k \cup Cand(Y_k)$ は，$Y_k$ に対し追加された頂点 $u_k$ と隣接しない $Cand(Y_k)$ 中の頂点が除去され，よって，$Y_k \cup Cand(Y_k)$ は単調減少列で，かつ除去された頂点との接続エッジが，クリークと候補の合併集合から外部への接続エッジへと変化する．すなわち，$y \in Y_k \subseteq Y_{k+1}$ から，$Y_{k+1}$ とその候補頂点集合の和集合 $Y_k \cup \{y_k\} \cup Cand(Y_k \cup \{y_k\})$ の外部に接続する $G_2$ のエッジ本数は単調に増加する．よって (2) を得る．

小節 3.2 で与えるアルゴリズムは下記に示す枝刈（$G_2$ **暫定評価枝刈**）を行う．

**事実 2.2** $G$ の極大クリーク $MC$ が $G$ のクリーク $C$ を含むとする．

$\delta_2 < est\text{-}E_2(C) \Rightarrow \delta_2 < est\text{-}E_2(C) \leq est\text{-}E_2(MC) = E_2(MC)$

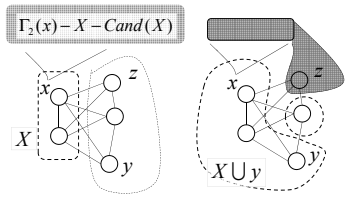


**図 2** クリークの $G_2$ 評価

### 2.3 変化前のパターン評価

統合グラフ $G$ のクリーク $X$ は，変化前のグラフ $G_1$ における独立集合であり，頂点集合として外部への接続度を測る尺度を本節で定める．変化後のパターン評価同様に，いくつかの代替案がありえる．例えば，$G_1$ における次数に関する固定順序を与え，頂点次数の平均次数で独立集合の外部への接続度を定める方法などもある．

こうした評価法もありえるが，本稿では，頂点の追加順序に依存しない形の定義を与えておく．基本的な考え方は，$E_2$ および $est\text{-}E_2$ 評価と同様で，現在の $G$ のクリーク $X$ の頂点および $X$ に将来追加される可能性のある候補頂点集合 $Cand(X)$ 中の頂点の $G_1$ における次数の最大値による暫定評価を行う．クリークの拡張 $X \subseteq X \cup \{u\}$ に対し，$X \cup \{u\} \cup Cand(X \cup \{u\}) \subseteq X \cup Cand(X)$ であるから，例えば次数の最大値により暫定評価を行えば，候補の追加に関して単調減少な暫定評価が得られる．

**定義 2.3** (1) $G$ の極大クリーク $MC$ に対し，$E_1(MC) = \max_{x \in MC} deg_1(x)$

(2) $G$ のクリーク $C$ に対し，$est\text{-}E_1(C) = \max_{x \in C \cup Cand(C)} deg_1(x)$

**事実 2.3** (1) $G$ のクリーク $C_1, C_2$ に対し，

$C_1 \subseteq C_2 \Rightarrow est\text{-}E_1(C_1) \geq est\text{-}E_1(C_2)$

(2) $G$ の極大クリーク $MC$ に対し，$est\text{-}E_1(MC) = E_1(MC)$

$G_1$ **暫定評価枝刈**： 特に，$C \subseteq MC$ かつ $\delta_1 > est\text{-}E_1(C) \Rightarrow \delta_1 > E_1(MC)$

上記 (1), (2) の性質を持つ評価法としては，例えば

$$est\text{-}E_1(X) = \sum_{x \in X \cup Cand(X)} deg_1(x)$$

など,（候補も含めた）頂点集合の包含関係に関して単調増加な評価関数ならなんでも良い．それらに対し，次節で述べる構造変化列挙の完全性は保たれる．

## 3. 構造変化パターン全列挙アルゴリズム

本節では，構造変化パターンを全列挙する方法について述べる．基本的には，極大クリーク全列挙手法[2] を，統合グラフの極大クリーク生成のために用いる．極大クリーク探索木を縦型に走査するバックトラック法である．ただし，極大クリークの重複枚挙を防ぎ，無駄な枝生成を抑制するためのシンプルで効果的な制御規則が組み込まれている．

本稿で必要なものは，統合グラフの極大クリークで，前節で述べた外部への接続制約を満たすものである．したがって，接続制約違反となる枝をさらにカットできる．探索木としては，統合グラフ極大クリーク生成のための探索木を「なぞり」ながら，その部分木のみを結果的に生成する．

### 3.1 極大クリーク全列挙

本小節では，論文[2] に基づいて極大クリーク全列挙法の要点を，探索木の言葉を用いて述べる．探索木の根は空集合であり，根から出発し，極大クリーク $MC$ に至るクリークの列 $X_0 = \phi, X_1, \cdots, X_n = MC$ をパスとして持つ．特に $MC$ は葉ノードを構成し，$X_{k+1}$ は，$X_k$ にその候補頂点 $x_k \in Cand(X_k) = \{y \in V \mid X_k \subseteq \Gamma(y)\}$ を加えた $X_k$ の子ノード $X_{k+1} = X_k \cup \{x_k\}$ として出現する．特に，$|MC|$ はパス長，また，$Cand(MC) = \phi$ であり，後者の条件は探索木におけるノード展開の一つの停止条件である．別の停止条件もあるが，これは，枚挙の重複を防ぐためために用いる（後述）．非極大クリーク $X_k$ は子ノード

を持つがその数は $|Cand(X_k)|$ の数とは一般に一致しない．この数が小さいほど探索効率は言うまでもなく向上し，そのために，論文[2] では下記に述べる制御を行う．なお，本稿では，説明のために，Left および Right 制御（候補）という言葉を用いる．

**Left 制御**： $X_k$ から子ノードを派生するために使われる候補頂点は，結果としてある順序がつき，これらを列 $x_{k_1}, \cdots x_{k_{d_k}}$ と記す．特に，$d_k \leq |Cand(X_k)|$ である．$X_k$ を含む全ての極大クリークは，どれかただ一つの $x_{k_j}$ を選択したときの子ノード $X_k \cup \{x_{k_j}\}$ に対する部分探索木の葉ノードとして出現すべく制御される．結果として，極大クリークの探索木における出現回数は丁度１である．この目的のために，$x_{k_j}$ を枝として選んだ場合，先行する全ての $x_{k_1}, \cdots, x_{k_{j-1}}$ を Left 候補として子ノード $X_{k+1} = X_k \cup \{x_{k_j}\}$ に記憶させる．$X_{k+1}$ の候補頂点集合 $Cand(X_{k+1})$ は，先行する頂点 $x_{k_j}$ を含んでも良いが，Left 候補であることから，$X_{k+1}$ の子ノードを張るための使用をブロックする．こうした Left 候補を親ノードから継承することにより，下記の性質が成立する．

非極大クリーク $Xq$ 以下の探索部分木では，$q$ より先行する Left $p$ に対し，$Xp$ を含む極大クリークは出現しない．

**Right 制御** ：複数の極大クリークが一つの非極大クリーク $X_k$ を包含する場合の処理である．この場合，$Cand(X_k)$ の要素 $y$ で，同じく $Cand(X_k)$ の別の要素 $z$ の隣接頂点として現れるものが存在する．$X_k \cup \{y\}$ を含む極大クリークは，上記のどれか一つの $z$ を含む極大クリークとして現れるので，$X_k \cup \{z\}$ 以下の部分木内で生成可能であり，$y$ の枝を張る必要はない．こうした不要の枝を刈るために，$Cand(X_k)$ 中，隣接頂点数が最大な候補頂点 $u = \mathrm{argmax}_{y \in Cand(X_k)} |Cand(X_k) \cap \Gamma(y)|$ を定める，$u$ と隣接した $y \in Cand(X_k)$ を Right 候補と定め，その枝はブロックされる．ここで，$u$ は Left であっても良いことに注意する．

**非極大クリーク $X_k$ における停止規則** ：$X_k$ においては，Left でも Right でもない候補頂点がある場合にのみ，その枝が生成される．逆に言えば，全ての $Cand(X_k)$ 内の候補頂点が Left もしくは Right になる場合は，$X_k$ を拡大して得られる極大クリークは別のパスで生成されることを意味し，よって，ただちにバックトラックして良い．

### 3.2 アルゴリズム

本稿で求めるパターン（頂点集合） $MC$ とは，

$MC$ は 統合グラフにおける極大クリーク
$E_1(MC) \geq \delta_1$ かつ $E_2(MC) \leq \delta_2$

の２つの性質を満たすものである．これを $\delta_1$ $\delta_2$-解と呼ぶ．統合グラフ $G$ における 解 $MC$ に含まれる非極大クリーク $X$ は，暫定評価が持つ単調性により常に

$est\text{-}E_1(X) \geq \delta_1$，かつ $est\text{-}E_2(X) \leq \delta_2$ （暫定評価テスト）

であり，また逆に，上記が満たされない場合は，$X$ を含むいかなる統合グラフの極大クリークは，$\delta_1$ $\delta_2$-解とはなりえない．よって，前節における極大クリーク全列挙における各探索ノードにおいて，暫定評価テストを行い，テストに失敗すれば直ちにバックトラックを行ってよい．図 3 に疑似コードを載せておく．

## 4. まとめと今後の展望

前節までで問題の所在と解法を述べてきた．実験については未だまとめきっておらず，発表時にまとめた上報告したい．実験において想定できる問題点，および今後の課題等について本節で述べる．

変化の後に，次数が急激に高くなる頂点同士が，統合グラフにおけるクリークを形成することは，特に，大事件の前後におけるグラフを扱う場合は十分に想定できるだろう．変化後の次数が高すぎるものは，顕著な頂点として抽出でき，そうした頂点間の結合関係は別枠で処理する形式をとることも十分妥当だと思われる．この立場から，下記では，変化後にあまりにも高すぎる次数の頂点は除外した上で考察を行うとする．

変化後の $G_2$ において，高次数のものと低次数のものが候補頂点として隣接している場合，高次数の候補頂点 $v$ と隣接した比較的多数の低次数頂点が Right 候補となる場合が問題となる．この場合，高次数の候補頂点が外部への接続エッジ数が大な場合，暫定評価により $v$ を含むクリークが枝刈されることも意味する．すなわち，$v$ よりも低次数の Right 候補を含むクリークの形成が阻まれる．この問題に対処する一つの方法は，統合グラフにおける極大クリーク生成性能はある程度落ちるが，

多数の候補頂点と隣接する候補頂点であっても，外部との接続が大きすぎるものは選択せず，他の候補による Right 候補決定を行う

規則の導入等，より細やかな制御が必要となる．このことにより，無駄な枝を生成しないという極大クリーク全列挙手法の優れた点をある程度キープし，かつ，本稿の研究目的を達成するための必要な経験則として導入の是非を検討する．

**procedure** StructuralChangeMain($G_1$, $G_2$, $\delta_1$, $\delta_2$):
  **[Input]** $G_1 = (V_1, \Gamma_1)$：変化前グラフ.
    $G_2 = (V_2, \Gamma_2)$：変化後グラフ.
    $\delta_1$：$G_1$ 側外部接続度下限閾値.
    $\delta_2$：$G_2$ 側外部接続度上限閾値.
  **[Output]**：すべての構造変化パターン族.
  **[Global Variables]** $G$：統合グラフ, $G_1$, $G_2$, $\delta_1$, $\delta_2$.
  **begin**
    $G \leftarrow (V = V_1 \cap V_2, \Gamma = \Gamma_1^c \cap \Gamma_2)$; // $G_1$ の補グラフと $G_2$ を統合
    StructuralChangeExpand($\emptyset$, $\emptyset$, $V$);
  **end**

---

**procedure** StructuralChangeExpand($Q$, $Left$, $Cand$):
  **if** $Cand = \emptyset$ **then**
    **if** $Left = \emptyset$ **then** output $Q$; // $Q$ は構造変化パターン
    return;
  **endif**
  $x \leftarrow$ a vertex $x$ in $Cand$ that maximizes $|Cand \cap \Gamma(x)|$;
  $Ext \leftarrow Cand \setminus \Gamma(x)$; // $Ext = Cand \setminus Right$
  **while** $Ext \neq \emptyset$ **do**
    $u \leftarrow$ a vertex in $Ext$;
    $NewQ \leftarrow Q \cup \{u\}$;
    $NewCand \leftarrow Cand \cap \Gamma(u)$;
    $NewLeft \leftarrow Left \cap \Gamma(u)$;
    **if** EstE1($NewQ$, $NewCand$) $\geq \delta_1 \wedge$
        EstE2($NewQ$, $NewCand$) $\leq \delta_2$ **then**
      StructuralChangeExpand($NewQ$, $NewLeft$, $NewCand$):
    **endif**
    $Cand \leftarrow Cand \setminus \{u\}$;
    $Left \leftarrow Left \cup \{u\}$;
    $Ext \leftarrow Ext \setminus \{u\}$;
  **endwhile**

---

EstE1($Q$, $Cand$):
  return $(\max_{x \in Q \cup Cand}\{|\Gamma_1(x)|\})$;

---

EstE2($Q$, $Cand$):

  return ( $\max_{x \in Q}\{\frac{|\Gamma_2(x) \setminus (Q \cup Cand)|}{|\Gamma_2(x)|}\}$ );

**図 3** 構造変化パターン全列挙アルゴリズム

また，上記の議論の前にそもそも実験的に確認すべき事項としては，グラフの規模や次数がそれほど高くない場合は，外部への接続条件を無視し，先に極大クリークを全列挙し，その事後処理として,「ハブ」的な頂点を除去し，接続条件を満たす非極大クリークを探査する方法も考えられる．これは，問題の規模と，Right 候補のより細かな制御の必要性とも関係するので，実験的にその必要性を明らかにしたい．

さらには，統合グラフの極大クリーク数がそもそも少数の場合は，極大疑似クリークの導入が考えられる．複雑な疑似クリークの場合，極大クリーク全列挙法がもつ枝生成制御が困難になることもありえる．$k$-Plex[1] として知られている疑似クリークは，今回の枝生成制御との親和性が高く，今後導入すべき有力な疑似クリーク法の一つである．

## 参 考 文 献

1) B. Wu and X. Pei, A Parallel Algorithm for Enumerating All the Maximal k-Plexes, Proc. of the PAKDD 2007 Workshops, LNAI-4819, pp. 476 - 483, 2007.
2) E. Tomita, A. Tanaka and H. Takahashi, The Worst-Case Time Complexity for Generating All Maximal Cliques and Computational Experiments, Theoretical Computer Science, 363(1), pp. 28 - 42, Elsevier, 2006.
3) M. E. J. Newman, Finding Community Structure in Networks Using the Eigenvectors of Matrices, Physical Revivew E, 74(3), 036104, American Physical Society, 2006.
4) T. Taniguchi and M. Haraguchi: Discovery of Hidden Correlations in a Local Transaction Database Based on Differences of Correlations, Engineering Application of Artificial Intelligence, 19(4), pp. 419 - 428, Elsevier, 2006.
5) A. Li, M. Haraguchi and Y. Okubo, Contrasting Correlations by an Efficient Double-Clique Condition, Proc. of the 7th International Conference on Machine Learning and Data Mining in Pattern Recognition - MLDM'11, Springer-LNAI 6871, pp. 469 - 483, 2011.
6) J. Kleinberg, Bursty and Hierarchical Structure in Streams, Data Mining and Knowledge Discovery, 7(4), pp. 373 - 397, Kluwer Academic Publishers, 2003.
7) P. Terlecki and K. Walczak, Efficient Discovery of Top-K Minimal Jumping Emerging Patterns, Proc. of the 6th International Conference on Rough Sets and Current Trends in Computing - RSCTC'08, Springer-LNAI 5306, pp. 438 - 447, 2008.
8) S. D. Bay, and M. J. Pazzani, Detecting Group Differences: Mining Contrast Sets, Data Mining and Knowledge Discovery, 5(3), pp. 213 - 246, Kluwer Academic Publishers, 2001.
9) 鶴田 哲章 ・原口 誠，モデュラリティの差異に基づくコントラスト法，人工知能学会全国大会 (第 25 回) 論文集，2G3-6, 2011.