

# 効率的アンサンブルに基づく不確実グラフにおける Motif-Role 抽出

内藤 綜志<sup>1,a)</sup> 伏見 卓恭<sup>1,b)</sup>

**概要:** 本論文では、不確実なエッジを持つグラフから Motif を拡張した概念である Motif-Role を抽出するという新たな問題を定式化する。はじめに、サンプリングされたグラフに存在する各ノードの Motif-Role をカウントする。次に、各ノードがどの Role として出現した頻度を基にノード間の類似性を計算する。そして、類似性に応じて全てのノードをクラスタリングする。不確実グラフから Motif-Role の抽出を高精度で行うには、多数回のグラフサンプリングをし、各サンプルグラフに対して、Role 数カウント、類似度計算、クラスタリングが必要になる。そこで、本研究ではサンプリングしたグラフをアンサンブルする手法を新たに提案し、既存のアンサンブル手法と結果の類似性・効率性の観点から比較評価する。各エッジに一樣な出現確率を付与した 4 つの実ネットワークを用いた実験から、提案法が最先端な Motif カウントアルゴリズムを使った既存アンサンブル法と比較して、同様の結果を高速に出力できることを示す。

## 1. はじめに

ネットワークサイエンスの分野において、Motif のカウントはそのグラフの特徴を理解する上で重要なタスクであり、また Milo らの研究 [1] から始まり長年多くの分野で研究されてきた [2], [3], [4], [5], [6]。Motif の概念を拡張する研究は盛んに行われており、有向 3 ノード Motif における構造同値性に基づき、各ノードの役割を定義した Motif-Role が提案され、それをを用いた分析結果が報告されている [7], [8]。図 1 は、グラフ同型に基づき有向 3 ノードサブグラフを 13 パターンに分類したものであり、Motif1~Motif13 として定義する。同様に、構造同値性に基づきサブグラフ内のノードを 30 パターンに分類したものであり、Role1~Role30 として定義する。この Motif に基づく Role を利用することで、例えば、Role13 として出現したノードは情報を発信することに特化した役割を担っており、また一方で Role24 として出現したノードは受け取った情報、他ノードに発信するといった役割を担っていると特徴づけることができる。このように、ノードごとの Motif における役割を抽出することでバイラルマーケティングにおいて重要なインフルエンサーなどの特定が可能になると期待できる。

また、ここ数年のネットワークサイエンスにおけるトレ

ンドとして不確実グラフ (Uncertain グラフ) の研究がある。不確実グラフは、エッジに対して出現確率が付与されたもので、道路ネットワークにおける通行止めや SNS でのユーザーのフォロー/フォロワー関係など、現実世界をより精緻に表現することが可能である。不確実グラフでの Motif や Role のカウントは、現実即したグラフの詳細な分析を行うことができるため、マーケティングをはじめとし、都市計画・タンパク質の分析といった幅広い分野での活用が期待される。しかし、エッジの存在が不確実なままでは、上述した Motif や Role の出現回数を確定的にカウントすることができないため、このような確率グラフに対しては、Motif や Role の出現回数の期待値を考えることになる。 $L$  本の不確実エッジを持つ不確実グラフに対して正確な期待値を求めるには、 $2^L$  個の起こりうるグラフを全て列挙し、それらに対して Motif または Role の数をカウントして、各グラフの生起確率による重み付きの平均値を計算する必要があるが、小規模なグラフであっても組み合わせ数  $2^L$  が非常に大きくなるため、一般に計算は困難である。そのため、一定数のサンプルを得るためのサンプリングを行い、それらをサンプル数で割ることで期待値の近似値を求めるという手法が一般的である。

不確実グラフにおける Motif のカウントに関する既存研究として、Ma らによる LINC アルゴリズム [9] がある。このアルゴリズムは、全サンプルグラフに対して 1 から Motif をカウントするのではなく、サンプルグラフ間の構造類似性に着目し、2 サンプルグラフ間の差分エッジおよびその

<sup>1</sup> 東京工科大学 コンピュータサイエンス学部  
School of Computer Science, Tokyo University of Technology  
a) c011833449@edu.teu.ac.jp  
b) fushimity@stf.teu.ac.jp

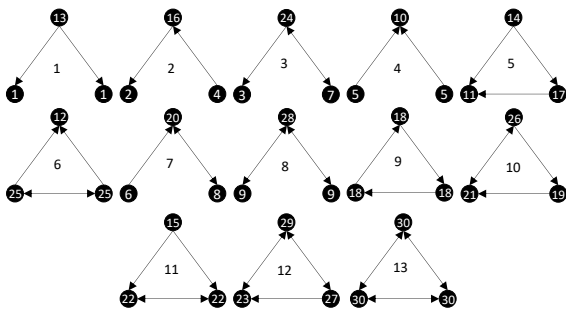


図 1 有向 3 ノード Motif における 30 Role

周辺のみ焦点を当てることで、効率的に Motif 数を更新する。LINC アルゴリズムは不確実グラフ上での Motif カウントアルゴリズムとして最先端のものであり、エッジの平均出現確率が高いグラフにおいてナイーブなサンプリング手法よりも高速に結果を出力できるが、エッジの出現確率が低いものを多く含むグラフや、エッジ数が多くなると実行速度が遅くなってしまいう課題が存在する。

本稿では Motif に基づく Role の抽出問題を定式化し、不確実グラフのサンプルに対する効率的なアンサンブルにより Role を抽出することを試みる。具体的には、サンプリングしたグラフをアンサンブルし、各エッジに全サンプル中での出現回数を重みとした重み付きグラフに対して Role を抽出する手法を提案する。そして、既存のベクトルアンサンブル法、類似度アンサンブル法と効率性や結果の類似性といった観点から比較し、提案法が最先端の手法である LINC アルゴリズムを導入したほか既存アンサンブル法と比較して、最も効率的であることを示す。

## 2. 関連研究

不確実グラフに関する研究は信頼性・クエリ・マイニングなど広いコンテキストで研究されてきており、Motif カウントを含む既存のグラフ分析手法を、不確実グラフに拡張することは重要なタスクである。Milo らによる先駆的な研究 [1] をきっかけに、モチーフのカウントの技術は何年にもわたって研究され、多くのアルゴリズムが多用途のために開発されてきた [2], [3], [4], [5], [6]。Wernicke は、ESU と呼ばれるハッシュベースのアルゴリズムを提案した。これは、すべてのサブグラフをハッシュテーブルに格納する必要性を回避し、同じサブグラフを 2 回カウントしないことでモチーフカウントの効率性を向上させた [2]。Itzhack らは、ターゲットノードをルートとする幅優先探索木をトラバースする効率的なアルゴリズムを提案した。これは、サブグラフ内のリンクの存在をビット文字列として表し、各サブグラフの同型性を調べることなくモチーフパターンを効率的に識別できる [3]。本研究においても、サンプルグラフからのモチーフカウントに、Itzhack らのアルゴリズムを採用する。Grochow と Kellis は、単一のモチー

フを検索するための非常に効率的なアルゴリズムを提案した [4]。このアルゴリズムでは、特定のグラフからターゲットモチーフへの部分的なマッピングを構築する。また、モチーフの複数カウントを回避するために symmetric-break と呼ばれる手法を導入し、実行時間の大幅な改善を実現した Ahmed らは、3 ノードと 4 ノードのモチーフを対象として、すべてのモチーフインスタンスを列挙せず、クリークやサイクルなどのいくつかのモチーフをカウントし、モチーフ間の遷移関係を使用して、他のすべてのモチーフを解析的に計算する並列アルゴリズムを提案した [5]。Pinar らは、見つかった各サブグラフの下部構造を識別して、それらをより小さなものに分割する分割統治アルゴリズムを提案した [6]。非常に効率的な手法であるが、有向ネットワークには適用できない。

上述した効率的なモチーフカウントの技術を活用し、グラフの頑健性 [10]、コミュニティ検索 [11]、ロールの抽出 [8]、グラフクラスタリング [12] などのさまざまな研究も行われている。また、McDonnell らは有向 3 ノードの構造同値性に注目することで Motif-Role を定義し、Motif の出現回数を表す行列を Role の出現回数を表す行列へと変換する変換行列を提案した [8]。本論文で対象とする Role は McDonnell らにより定義された Motif-Role と同じものであるが、カウンティングに使用するアルゴリズムは McDonnell らのものではなく、前述した Itzhack らによるものを用いる。

不確実グラフに対する Motif カウントの研究はまだ十分にされておらず、以下に主要なものをあげる。Tran らは、偽のエッジと欠損エッジの存在を考慮して、Motif 数の不偏推定量を計算する方法を提案したが、すべてのエッジが一樣な出現確率であることを想定した手法であり、非一樣な確率には適用できない [13]。Ma らは、サンプリングに基づいて、モチーフ数の平均、分散、確率分布などの基本的な統計を求める 2 つのアルゴリズムを提案した [9]。1 つ目は、PGS と呼ばれる単純なサンプリング方法で、不確実グラフから多数の実現グラフをサンプリングし、各サンプルグラフから単一のモチーフのインスタンスをカウントする。さらに、Hoeffding の不等式に基づいてモチーフ数の平均を正確に推定するのに十分なサンプル数について議論している。2 つ目は、LINC と呼ばれるより効率的な方法で、サンプルグラフ間の構造的類似性を利用し、連続するサンプル間のエッジの違いのみを調べ、モチーフの頻度を更新する。LINC は、全く同じサンプルグラフを使用した場合、PGS と同じ結果を出力するが、PGS よりもはるかに高速に動作するアルゴリズムである。本研究では、LINC アルゴリズムを最先端の技術と見なす。

### 3. 問題設定

#### 3.1 Motif-role 抽出

はじめに、エッジの出現確率がすべて1である決定的グラフ  $G = (V, E)$  から Motif-Role を抽出する問題を定式化する。ここで、 $V$  はノード集合、 $E$  はエッジ集合、 $N = |V|$  はノード数、 $L = |E|$  はエッジ数である。McDonnell らの研究 [8] に従い、Motif における構造的同値性に基づき Role を定義する。有向3ノード Motif においては、図1に示す30種類の Role が存在する。本研究における Motif-Role

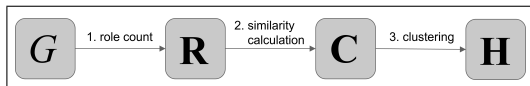


図2 Motif-Role の抽出手順

抽出は、1) ノードごとの Role ベクトルの構築、2) 全ノードペア間の Role ベクトル間の類似度計算、3) Role ベクトル間の類似度に基づきクラスタリングの3ステップからなる(図. 2)。手順1のノードごとの Role ベクトルの構築では、各ノード  $u$  に対して  $R$  個の Role の数をカウントし、その出現頻度を並べた  $R$  次元ベクトル  $\mathbf{r}_u$  を構築する。 $\mathbf{r}_u$  における  $j$  番目の要素は、ノード  $u$  が  $j$  の Role として出現した回数を表している。有向3ノード Motif の場合、図1に示さように  $R = 30$  である。全  $N$  ノードの Role ベクトルを並べた行列を  $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_N]^T$  と表記する。ここで  $\mathbf{r}^T$  は  $\mathbf{r}$  の転置を表す。手順2では、コサイン類似度  $c(u, v) = \frac{\mathbf{r}_u^T \mathbf{r}_v}{\|\mathbf{r}_u\| \|\mathbf{r}_v\|}$  を用い、全てのノードペアの類似度を計算する。全  $N \times N$  のノードペアの類似度を類似度行列  $\mathbf{C} = [c(u, v)]_{u \in V, v \in V}$  とする。手順3では、全てのノードを  $k$ -medoids クラスタリングの貪欲法 [14] により  $K$  個のクラスタに分類する。すなわち、もしノード  $u$  がクラスタ  $k$  に所属している場合、 $h_{u,k} = 1$  そうでない場合は0とする。 $h_{u,k} \in \{0, 1\}$  はノード  $u$  がクラスタ  $k$  に所属している情報、 $\mathbf{h}_u = [h_{u,k}]_{k=1}^K$  は所属ベクトルとなり、 $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N]^T$  が出力となる。結果としては、類似の Role ベクトルを有するノードクラスタが得られる。この一連のプロセスを Role 抽出と定義する。

#### 3.2 不確実グラフでの Motif-Role 抽出

次に、不確実グラフ  $\mathcal{G} = (G, p)$  での Motif-Role 抽出を定式化する。不確実グラフ  $\mathcal{G} = (G, p)$  はバックボーングラフ  $G = (V, E)$  と、各エッジの存在確率  $p: E \rightarrow (0, 1]$  により定義される。 $\mathcal{G} = (G, p)$  に対して、上で定義した Role 抽出問題を厳密に解くには、全ての実現可能グラフ  $G = \{G_i = (V, E_i); E_i \subseteq E\}$  に対して上述の3つの手順を行い、それらの結果を各実現可能グラフ  $G_i$  の生起確率  $\Pr[G_i]$  を考慮してアンサンブルする必要がある。不確実グ

ラフの関連研究に倣い、実現可能グラフ  $G_i$  の生起確率をすべてのエッジに対する独立ベルヌーイ試行に基づき計算する：

$$\Pr[G_i] = \prod_{e \in E_i} p(e) \prod_{e \in E \setminus E_i} (1 - p(e)).$$

そして、以下のように各実現可能グラフに対する Role 抽出結果をアンサンブルする：

$$\mathcal{H} = \Phi_{G \in \mathcal{G}}(\mathbf{H}_G; \Pr[G]).$$

ここで、 $\Phi$  はアンサンブルの演算子であり、各決定グラフ  $G$  のクラスタリング結果  $\mathbf{H}_G$  が生起確率  $\Pr[G]$  の重みを考慮してアンサンブルされていることを表す。不確実エッジの数  $L = |E|$  の不確実グラフに対して厳密なアンサンブル結果を得るには、実現グラフの数  $|\mathcal{G}| = 2^L$  だけサンプリングおよびアンサンブルが必要になり、小規模なグラフに対しても計算は困難であるため、サンプリングによる近似が一般的に採用される。

### 4. 提案手法

本論文では、不確実グラフから実現可能グラフをサンプリングし、それらを合成した重み付きグラフに対して Role カウント、類似度計算、およびクラスタリングをするグラフアンサンブル法を提案する。図3は、文献 [15] で提案

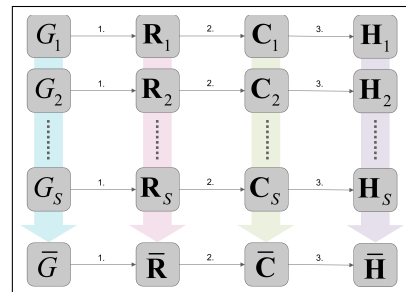


図3 4つのアンサンブル法

したベクトルアンサンブル法と類似度アンサンブル法に加え、本稿で新たに提案するグラフアンサンブル法および今後の課題として残るクラスタアンサンブル法の4つの手法を模式的に描いたものである。

ここでは、新たに提案するグラフアンサンブル法についてのみ説明する。グラフアンサンブル法はサンプルグラフ  $\{G_1, \dots, G_S\}$ 、 $G_s = (V, E_s)$ 、 $E_s \subseteq E$  に対しアンサンブルを行い、アンサンブルグラフ  $\bar{G}$  を生成する：

$$\bar{G} = \Phi_{G \in \mathcal{G}}(G; \Pr[G]) \simeq \Phi_{s=1}^S(G_s; 1/S) = \bar{G}.$$

$\bar{G} = (V, \bar{E}, w)$  は  $S$  回のサンプルの中でエッジ  $e$  が出現する確率  $w(e) = \sum_{s=1}^S \delta(e \in E_s) / S$  を重みとした重み付きグラフである。 $\delta(cond)$  は条件式  $cond$  が True の場合1を

返し, False の場合 0 を返す Boolean 関数である. 本手法は, アンサンブルグラフ  $G$  において, 各ノードがどの Role として出現したかを, 重み  $w(e)$  を考慮してカウントし, 出現回数を要素とした Role ベクトル群 (行列)  $\bar{\mathbf{R}}$  を構築する. そして, ベクトル間の類似度を要素とした行列  $\bar{\mathbf{C}}$  に基づいて, 各ノードをクラスターに分類することで,  $\bar{\mathbf{H}}$  を出力する. 本手法では  $L$  個のエッジを持った  $S$  個のグラフのアンサンブルし, アンサンブルグラフに対し一度だけ全  $N$  ノードの Role 数をカウントする. したがって,  $\bar{d}$  を平均度数とすると, 本手法の計算量は  $k$  ノード Motif において,  $O(SL + Nd^{\bar{d}(k-1)})$  となる.

## 5. 比較手法

文献 [15] で提案したベクトルアンサンブル法と類似度アンサンブル法について説明する.

### 5.1 Role-vector ensemble

ロールベクトルアンサンブル法 (ベクトルアンサンブル法) は, サンプルグラフ  $G_s$  から得られたロールベクトル  $\{\mathbf{R}_1, \dots, \mathbf{R}_S\}$  を平均することで, アンサンブル・ロールベクトル  $\bar{\mathbf{R}}$  を生成する:

$$\bar{\mathbf{R}} = \Phi_{G \in \mathcal{G}}(\mathbf{R}_G; \Pr[G]) \simeq \frac{1}{S} \sum_{s=1}^S \mathbf{R}_s = \bar{\mathbf{R}}.$$

そして, 得られたアンサンブル・ロールベクトル  $\bar{\mathbf{R}}$  からコサイン類似度  $\bar{\mathbf{C}}$  を計算し, 類似度行列をもとに各ノードをクラスターに振り分け,  $\bar{\mathbf{H}}$  を出力する. 各サンプルグラフ  $G_s$  からロールベクトル  $\mathbf{R}_s$  を構築する際, 最先端技術である LINC アルゴリズム [9] に従い, 2つの連続するサンプルグラフにおけるエッジ集合  $E_s$  と  $E_{s+1}$  間の差分  $D_{s,s+1} = (E_s \setminus E_{s+1}) \cup (E_{s+1} \setminus E_s)$  に着目し, 状態が変化したエッジ  $e \in D_{s,s+1}$  に関する Role の出現数のみを更新する.  $p$  をエッジの平均出現確率とすると, 状態変化するエッジ数の期待値は  $2L(p - p^2)$  となるため,  $p = 0.1$  や  $p = 0.9$  のとき効果を発揮する. このようにして,  $S$  個の  $(N \times R)$  ロールベクトル  $\{\mathbf{R}_1, \dots, \mathbf{R}_S\}$  を効率的に求め, 平均を計算する. したがって,  $\bar{m}$  を各エッジに関するモチーフインスタンスの平均数とすると, ベクトルアンサンブル法の計算量は  $O(S(NR + L(p - p^2)\bar{m}))$  となる.

### 5.2 類似度アンサンブル

類似度アンサンブル法は, ロールベクトル群  $\{\mathbf{R}_1, \dots, \mathbf{R}_S\}$  から計算された類似度行列  $\{\mathbf{C}_1, \dots, \mathbf{C}_S\}$  の平均を求め, アンサンブル類似度行列  $\bar{\mathbf{C}}$  を生成する:

$$\bar{\mathbf{C}} = \Phi_{G \in \mathcal{G}}(\mathbf{C}_G; \Pr[G]) \simeq \frac{1}{S} \sum_{s=1}^S \mathbf{C}_s = \bar{\mathbf{C}}.$$

アンサンブル・類似度行列  $\bar{\mathbf{C}}$  に基づき各ノードをクラス

ターに振り分け,  $\bar{\mathbf{H}}$  を出力する. ベクトルアンサンブル法と同様に, LINC アルゴリズムに基づき Role 数をカウントする. このようにして,  $S$  個の  $(N \times N)$  の類似度行列を求め, そして平均を計算するため, 本手法の主たる計算量は  $O(SN^2)$  となる \*1.

### 5.3 クラスタアンサンブル法

クラスタアンサンブル法は, クラスタリング結果をアンサンブルし, 所属クラスタ行列  $\{\mathbf{H}_1, \dots, \mathbf{H}_S\}$  を生成する:

$$\bar{\mathbf{H}} = \Phi_{G \in \mathcal{G}}(\mathbf{H}_G; \Pr[G]) \simeq \Phi_{s=1}^S(\mathbf{H}_s; 1/S) = \bar{\mathbf{H}}.$$

クラスタリング結果をアンサンブル手法はまだ確立されていないため, 今回は触れない.

## 6. 評価実験

### 6.1 実験データ

評価実験では, 後述する 3つの現実世界の有向グラフを用いた有向 3 ノードモチーフをもとにした Motif-Role の抽出を行い, 本研究の提案手法であるグラフアンサンブルの効率性を評価する. 実験に使うグラフを表 1 に示す. これらのグラフ

表 1 使用する実有向グラフ

データの種類	ノード数	エッジ数
Celegans(線虫の神経回路網)	131	764
Gnutella(P2P ネットワーク)	10,876	39,994
Enron (従業員間のメール送受信) [16]	19,603	210,950

の各エッジに対し, 一様な出現確率  $p(e) = p \in [0.5, 0.9]$  を付与した. サンプル数は  $S \in \{10^1, 10^2, 10^3, 10^4\}$  とし, クラスタ数は  $K = 10$  とした \*2. サンプル数やクラスタ数の違いにより, 結果の類似性や実行時間に影響があるか実験的に検証する. 紙面の都合上, クラスタ数  $k = 10$  の結果のみを示す.

### 6.2 ロールベクトルの類似性評価

はじめに, ロールベクトルの類似性を, 各ノードのコサイン類似度の平均値で評価する. ノード  $v$  について,  $\mathbf{r}_v^{(gra)}$  と  $\mathbf{r}_v^{(vec)}$  をグラフアンサンブル法とベクトルアンサンブル法により求められたロールベクトルとし, コサイン類似度  $\cos(v) = \langle \mathbf{r}_v^{(gra)}, \mathbf{r}_v^{(vec)} \rangle / \|\mathbf{r}_v^{(gra)}\| \|\mathbf{r}_v^{(vec)}\|$  を求める. そして, 全ノードのコサイン類似度の平均値  $\frac{1}{N} \sum_{v \in V} \cos(v)$  を求める. 図 4 は, 横軸にサンプル数  $S$ , 縦軸に平均コサイン類似度をプロットした. これらの図から, エッジの出現確率が高いほど, コサイン類似度は高くなることがわかり, グラフアンサンブルとベクトルアンサンブルは非常に類似した結果を出力することが確認できた.

\*1 LINC アルゴリズムに基づき Role 数をカウントするが, 類似度行列のアンサンブルにかかる計算量が支配的となる

\*2 Celegans のみはグラフサイズが小さいため  $K = 5$  とした.

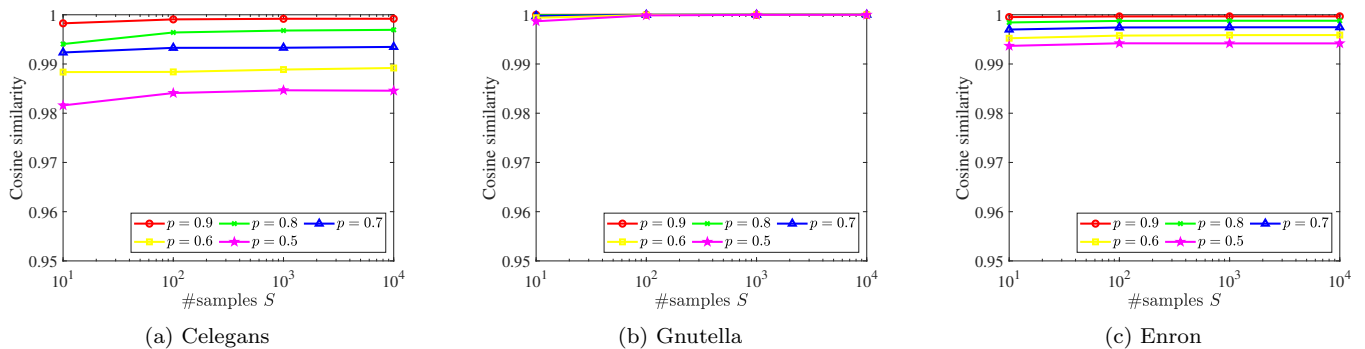


図 4 各アンサンブル法のロールベクトル間の平均コサイン類似度

### 6.3 クラスタリング結果の類似性評価

つぎに、Motif-Role 抽出法の最終的な出力となるクラスタリング結果の類似性を評価する。本研究では、 $k$ -medoids クラスタリングの貪欲法を採用しているため、同じ類似度行列を入力した場合、常に同じ結果を出力し、実行するたびに結果が変わるランダム性はない。評価には、クスタ間の類似度を計測する正規化相互情報量 (NMI) [17] を用いる。図 5 から、Celegans はグラフが小さいため安定した結果は得られなかったものの、その他のグラフにおいてはサンプル数、出現確率の値が高いほど、類似した結果を出力することが確認できた。

### 6.4 効率性評価

最後に、各アンサンブル法を実行時間の観点で評価をする。図 6 は各手法がクラスタリング結果を出力するまでの実行時間を示す。横軸はサンプル数、縦軸は対数スケールで表した実行時間である。これらのグラフから、以下のことが確認できる。1) グラフアンサンブル法は、他の手法より圧倒的に高速である。2) グラフアンサンブル法はサンプル数やエッジの出現確率による影響をほとんど受けない。3) ベクトルアンサンブル法はサンプル数やエッジの出現確率に多少の影響を受ける。4) 類似度アンサンブル法はサンプル数やエッジの出現確率に線形的に影響を受ける。

類似性評価および効率性評価の結果から、グラフアンサンブル法は最先端の手法である LINC アルゴリズムを導入した他のアンサンブル法と非常に類似した結果を非常に高速に出力できると結論づける。

## 7. おわりに

本稿では、既存タスクである Motif カウントを拡張した Role カウントを不確実グラフに対して行う問題を新たに定式化し、サンプリングしたグラフをアンサンブルするグラフアンサンブル法を提案した。実験では、4つの実ネットワークに一樣なエッジ確率を付与したのを用い、既存のアンサンブル手法と比較して、結果の類似性および計算時間の観点で評価した。結果として、本稿で提案したグラフアンサンブル法が最先端手法である LINC を導入した既

存アンサンブル法（ベクトルアンサンブル法、類似度アンサンブル法）と比べて、実行時間においてははるかに効率的であることが確認できた。また、いずれの手法も出力結果は類似することも確認できた。以上のことより、グラフアンサンブル法が本稿で扱う問題に最も適していると結論づけた。

今後の課題として、Role 数の正確な期待値を求め、各アンサンブル手法における Role 数の誤差評価、Hoeffding の不等式を用いた適切なサンプリング数の決定、クラスタリング結果のより詳細な分析などがあげられる。

謝辞 本研究は、JSPS 科研費 (No.20K11940) (No.19K20417) の助成を受けたものである。

### 参考文献

- [1] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: simple building blocks of complex networks., *Science (New York, N. Y.)*, Vol. 298, No. 5594, pp. 824–827 (online), DOI: 10.1126/science.298.5594.824 (2002).
- [2] Wernicke, S.: A Faster Algorithm for Detecting Network Motifs, *Proceedings of the 5th International Conference on Algorithms in Bioinformatics, WABI'05*, Berlin, Heidelberg, Springer-Verlag, pp. 165–177 (2005).
- [3] Itzhack, R., Mogilevski, Y. and Louzoum, Y.: An optimal algorithm for counting network motifs, *Physica A: Statistical Mechanics and its Applications*, Vol. 381, pp. 482–490 (online), DOI: <https://doi.org/10.1016/j.physa.2007.02.102> (2007).
- [4] Grochow, J. A. and Kellis, M.: Network Motif Discovery Using Subgraph Enumeration and Symmetry-Breaking, *Proceedings of the 11th Annual International Conference on Research in Computational Molecular Biology, RECOMB'07*, Berlin, Heidelberg, Springer-Verlag, pp. 92–106 (2007).
- [5] Ahmed, N. K., Neville, J., Rossi, R. A. and Duffield, N.: Efficient Graphlet Counting for Large Networks, *2015 IEEE International Conference on Data Mining*, pp. 1–10 (online), DOI: 10.1109/ICDM.2015.141 (2015).
- [6] Pinar, A., Seshadhri, C. and Vishal, V.: ESCAPE: Efficiently Counting All 5-Vertex Subgraphs, *Proceedings of the 26th International Conference on World Wide Web, WWW '17*, Republic and Canton of Geneva, CHE, pp. 1431–1440 (online), DOI: 10.1145/3038912.3052597 (2017).
- [7] Ohnishi, T., Takayasu, H. and Takayasu, M.: Network



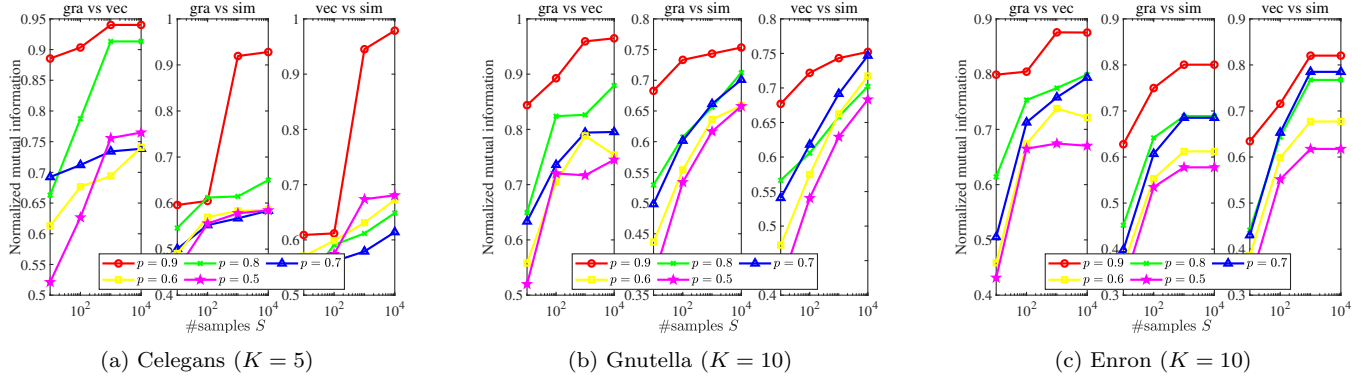


図 5 各アンサンブル法の  $k$ -medoids クラスタリングの結果間の正規化相互情報量. 左: グラフアンサンブル法 vs ベクトルアンサンブル法; 中央: グラフアンサンブル法 vs 類似度アンサンブル法; 右: ベクトルアンサンブル法 vs 類似度アンサンブル法;

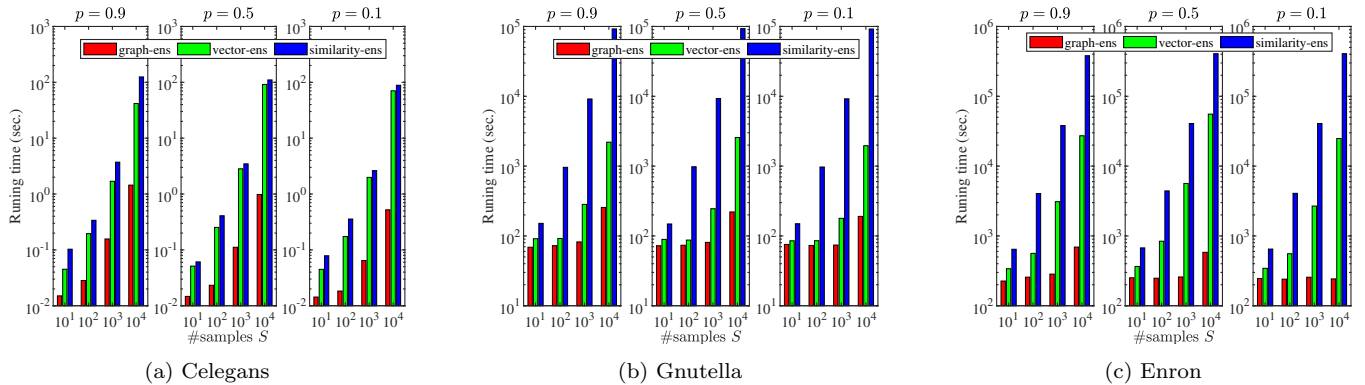


図 6 実行時間

motifs in an inter-firm network, *Journal of Economic Interaction and Coordination*, Vol. 5, No. 2, pp. 171–180 (2010).

[8] McDonnell, M. D., Yaveroglu, O. N., Schmerl, B. A., Iannella, N. and Ward, L. M.: Motif-Role-Fingerprints: The Building-Blocks of Motifs, Clustering-Coefficients and Transitivity in Directed Networks, *PLOS ONE*, Vol. 9, No. 12, pp. 1–25 (online), DOI: 10.1371/journal.pone.0114503 (2014).

[9] Ma, C., Cheng, R., Lakshmanan, L. V. S., Grubemann, T., Fang, Y. and Li, X.: LINC: A Motif Counting Algorithm for Uncertain Graphs, *Proc. VLDB Endow.*, Vol. 13, No. 2, p. 155–168 (online), DOI: 10.14778/3364324.3364330 (2019).

[10] Kaluza, P., Vingron, M. and Mikhailov, A. S.: Self-correcting networks: Function, robustness, and motif distributions in biological signal processing, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 18, No. 02, pp. 026113:1–026113:17 (2008).

[11] Huang, X., Cheng, H., Qin, L., Tian, W. and Yu, J. X.: Querying  $K$ -Truss Community in Large and Dynamic Graphs, *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14*, New York, NY, USA, Association for Computing Machinery, pp. 1311–1322 (online), DOI: 10.1145/2588555.2610495 (2014).

[12] Yin, H., Benson, A. R., Leskovec, J. and Gleich, D. F.: Local Higher-Order Graph Clustering, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, New York, NY, USA, Association for Computing Machinery, pp. 555–564 (online), DOI: 10.1145/3097983.3098069

(2017).

[13] Tran, N., Choi, K. P. and Zhang, L.: Counting motifs in the human interactome, *Nature communications*, Vol. 4, p. 2241 (online), DOI: 10.1038/ncomms3241 (2013).

[14] Nemhauser, G. L., Wolsey, L. A. and Fisher, M. L.: An Analysis of Approximations for Maximizing Submodular Set Functions, *Mathematical Programming*, Vol. 14, pp. 265–294 (1978).

[15] 内藤綜志, 伏見卓恭: Uncertain グラフに対する効率的なサンプリングに基づく motif-role ベクトルのクラスタリング, 人工知能学会研究会資料知識ベースシステム研究会, Vol. 123, pp. 45–50 (2021).

[16] Leskovec, J. and Krevl, A.: SNAP Datasets: Stanford Large Network Dataset Collection, <http://snap.stanford.edu/data> (2014).

[17] Kvålseth, T. O.: On Normalized Mutual Information: Measure Derivations and Properties, *Entropy*, Vol. 19, No. 11 (online), DOI: 10.3390/e19110631 (2017).