

# 日本全国の商流データを用いた GNN に基づく中核企業の識別

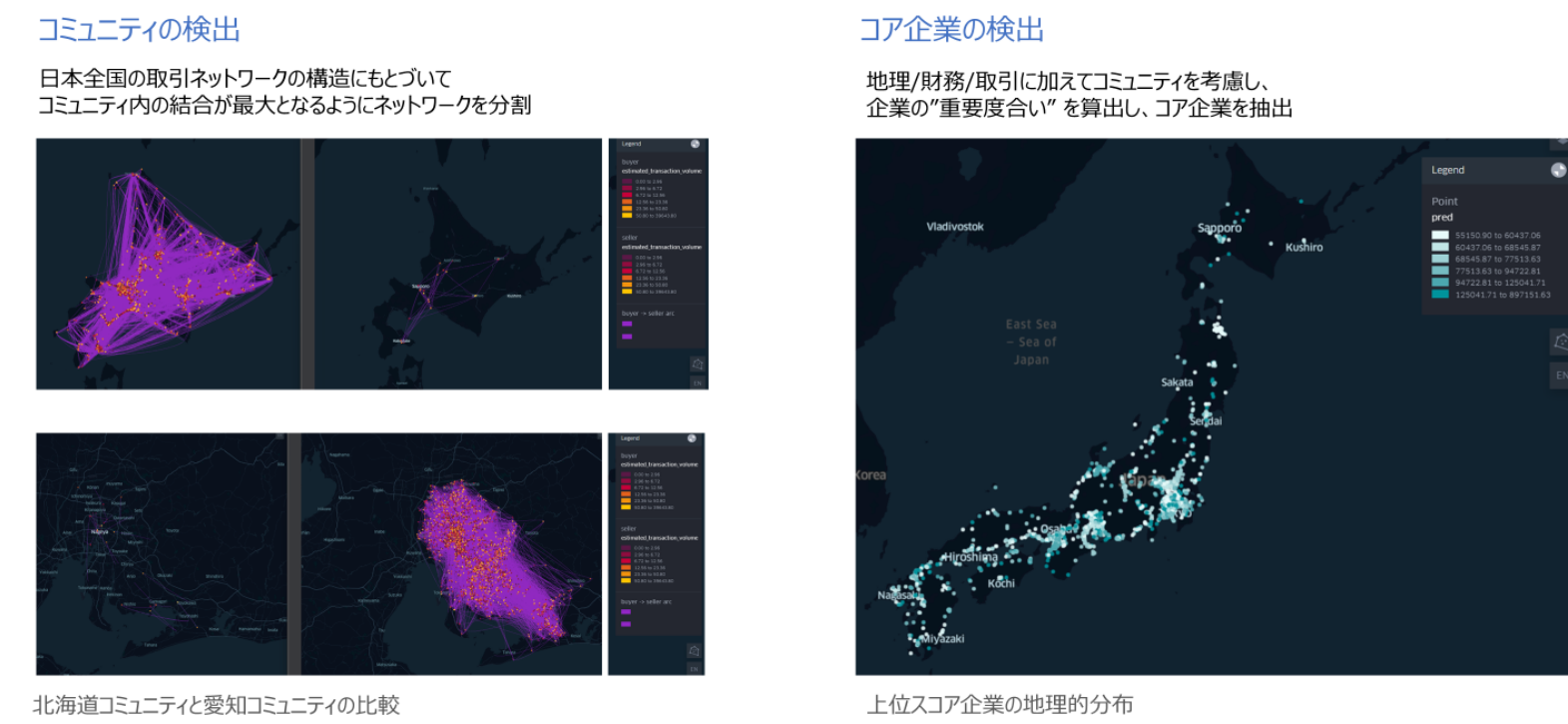
鹿野 英明, 小川 芳樹, Soohyun JOO, 関本 義秀

## 研究の背景

レジリエントなサプライチェーンを構築するためには、商流における中核企業を特定することが重要

### 本研究 | Core Company Identification とは

ボトルネック企業とコネクタ企業を包括する概念として新たに Core Company を定め、Graph Neural Network の枠組みを用いて企業間取引において重要な役割を果たす企業を検出すること  
具体的に実現を目指すこと



### 先行研究 | ボトルネック企業の検出

企業間取引データと企業概要データに基づき、テーブルデータに対する教師有り学習によってボトルネック企業を特定

サプライチェーンにおけるボトルネック企業の抽出モデル [Ogawa et al. 2022]

#### 企業間取引データ

GIMT (2019年) から得られる約84万社による企業間取引データ

#### ボトルネック企業データ

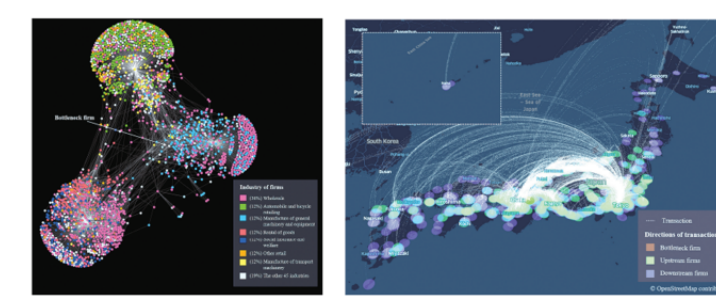
日経新聞によるアンケート調査 (2012) に基づき 28 社を選択

#### 学習に活用する特徴量

- 取引データと概要情報より以下を作成
- 企業に関する基本属性(売上、資本、従業員数)
- サプライチェーンネットワークの中心性
- 局所ネットワークの構造に関する指標

#### 分類モデル

- 複数の機械学習モデルで学習および分類
- ロジスティック回帰
- サポートベクターマシン
- ランダムフォレスト

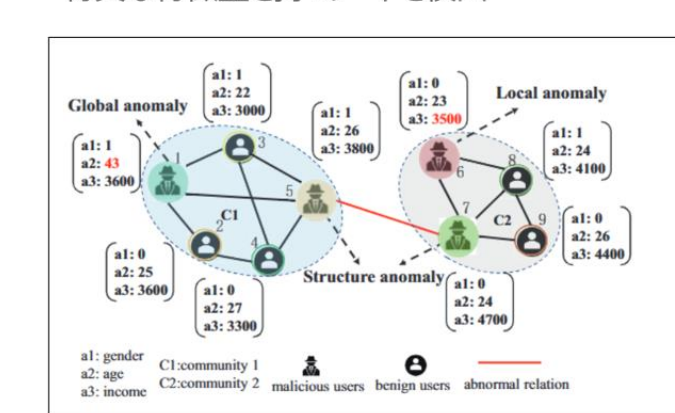


### 先行研究 | Graph Anomaly Detection

最近 Graph Neural Network を用いて、特異ノードを検出する研究が進展  
ノードの特徴だけでなくグラフの連結状態やコミュニティを考慮して特異を検出

#### コミュニティに着目した特異ノード検出

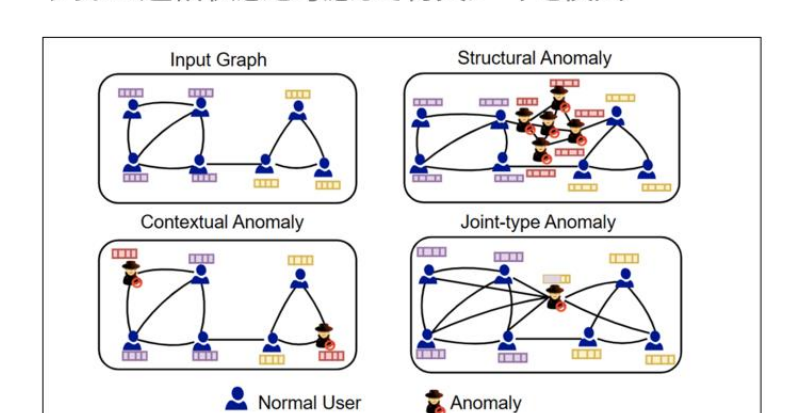
グラフのコミュニティ情報を考慮して、特異な特徴量を持つノードを検出



Luo et al. (WSDM'22)

#### グラフ構造に着目した特異ノード検出

特異な特徴量を持つノードだけでなく、グラフの連結状態を考慮して特異ノードを検出



Roy et al. (WSDM'24)

## 研究の貢献

商流をグラフとみなし日本全国の企業間取引データを用いて GNN に基づき商流の中核企業を識別

### Graph Neural Network | Core Company Identification の枠組み

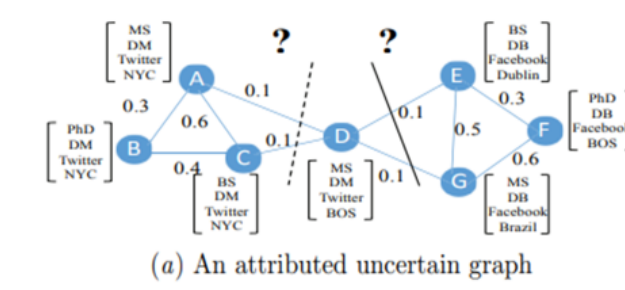
企業間取引をグラフの問題として捉え、コミュニティおよびコア企業の検出を行う

#### GNN の問題としての定式化

グラフとしての企業間取引  
企業間取引を属性グラフ  $G = (V, E, X)$  と考える

- ノード  $V$  を企業とみなし、以下のように考える  
 $V = \{v_1, v_2, \dots, v_n\}, |V| = n$
- エッジ  $E$  を企業間取引とみなし以下のように考える  
 $|E| = m$
- ノードに紐づく属性  $X$  を企業の特徴とみなし、以下のように考える  
 $X \in \mathcal{R}^{n \times d}$

#### 属性グラフの例



#### 主なモジュール

##### コミュニティの検出

METIS を用いてグラフを分割  
これにより分割されたグラフのサイズはほぼ等しくなり、サブグラフ間の辺の数は最小限になるように最適化される  
つまり、コミュニティ内の取引を最大、かつ、コミュニティ間の取引を最小にするように分割する

##### コア企業の検出

以下の値を各企業における目的変数として学習し、その予測値を重要度を表すスコアとして扱う

$$\begin{aligned} \mu_{\text{取引社数}} &= \sum_{j \in I} (1[\text{受注取引}_{i,j}]) + \sum_{j \in I} (1[\text{発注取引}_{i,j}]) \\ \mu_{\text{取引額}} &= \sum_{j \in I} (\text{受注取引額}_{i,j}) + \sum_{j \in I} (\text{発注取引額}_{i,j}) \\ \mu_{\text{コア}} &= \{\sum_{j \in I} (1[\text{受注取引}_{i,j}]) + \sum_{j \in I} (1[\text{発注取引}_{i,j}])\} \\ &\quad \times \{\log(\sum_{j \in I} (\text{受注取引額}_{i,j})) + \log(\sum_{j \in I} (\text{発注取引額}_{i,j}))\} \\ &\quad \times \{\sum_{j \in I} (1[\text{異なるコミュニティ}_{i,j}])\} \end{aligned}$$

### データ | 活用データとデータ加工

NIHACHI データとZAIMUH データをもとに企業ごとに特徴量を作成し GNN モデルの学習に活用

#### 活用するデータ

##### NIHACHI データ

2021年の日本全国の年間取引データ4,528,642 件を活用  
企業数は 620,068 社

受注社企業CD	発注社企業CD	取引高推計値
ABC001	XYZ001	291.24
ABC002	XYZ001	34.79
...	...	...

##### ZAIMUH データ

ZAIMUH に記載されている 802,017社の以下のデータを活用  
従業員数/総資本/自己資本/売上高/経常収入/営業支出/  
販管費/買入債務前渡金増減/棚卸資産増減/資本金

#### モデルの入出力

##### モデル入力

企業コード	結合特徴量	財務特徴量	コミュニティ特徴量
ABC001	(1, 0, 49, ...)	(10000, 132, ...)	0
ABC002	(3, 1, 22, ...)	(3000, 212, ...)	24
...	...	...	...

##### モデル出力

企業コード	スコア
ABC001	4.5
ABC002	5.9
...	...

## 計算手法

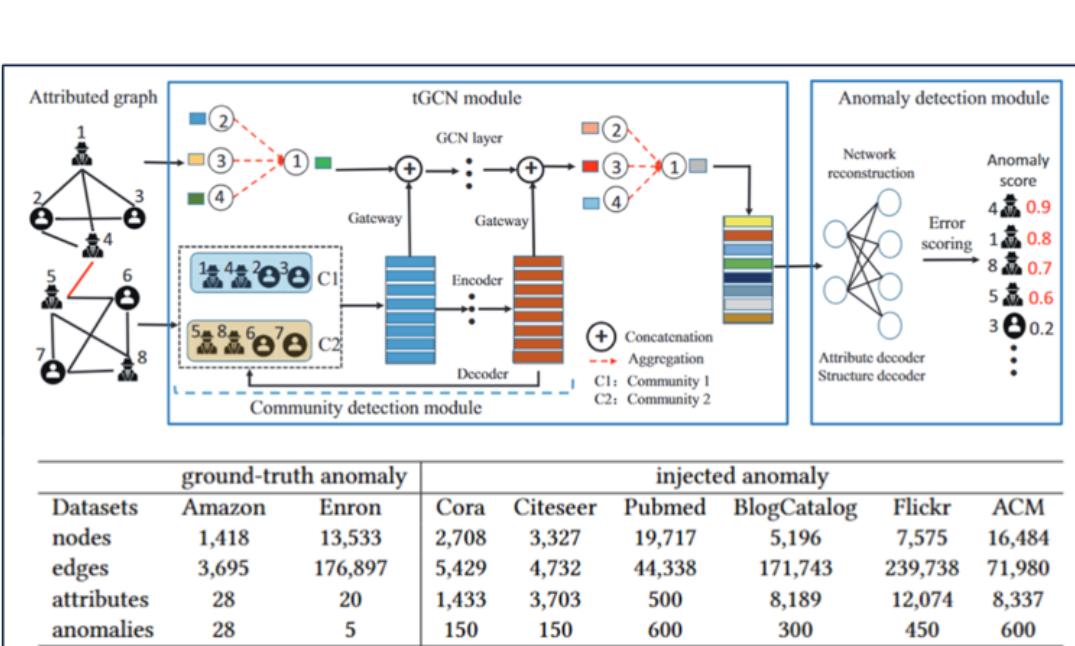
商流からコミュニティを検出した後に取引高/取引社数/取引コミュニティが多いほど高いスコアを付与

### Graph Neural Network | コミュニティの検出

日本全国を計算の対象とすると先行研究の手法では計算量的に困難  
そこで、ボトルネックやコネクタ企業がネットワークの中心にあることに着目して、コミュニティを検出

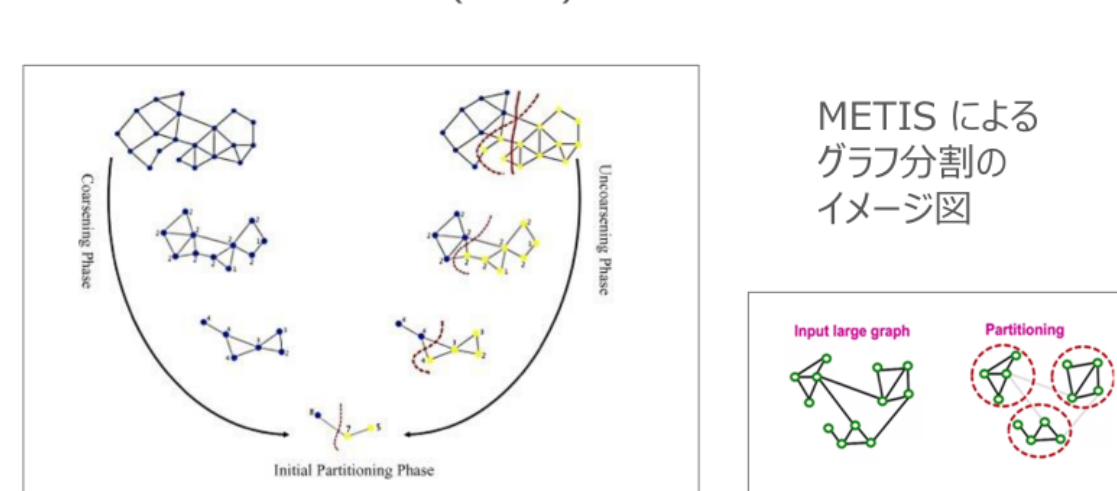
#### 日本全国規模のデータを扱う際の課題 - 計算量

先行研究における手法では、隣接行列に基づく計算によってコミュニティを算出し、しかしながら、この手法では計算コストが高く、日本全国の取引である NIHACHI を用いた計算は困難



#### METIS を用いたグラフ分割

そこで、ボトルネック企業やコネクタ企業が取引ネットワークの中心に存在する企業であることに着目し、最小コミュニティ間-最大コミュニティ内となるように、取引ネットワークを METIS (1998) を用いて分割



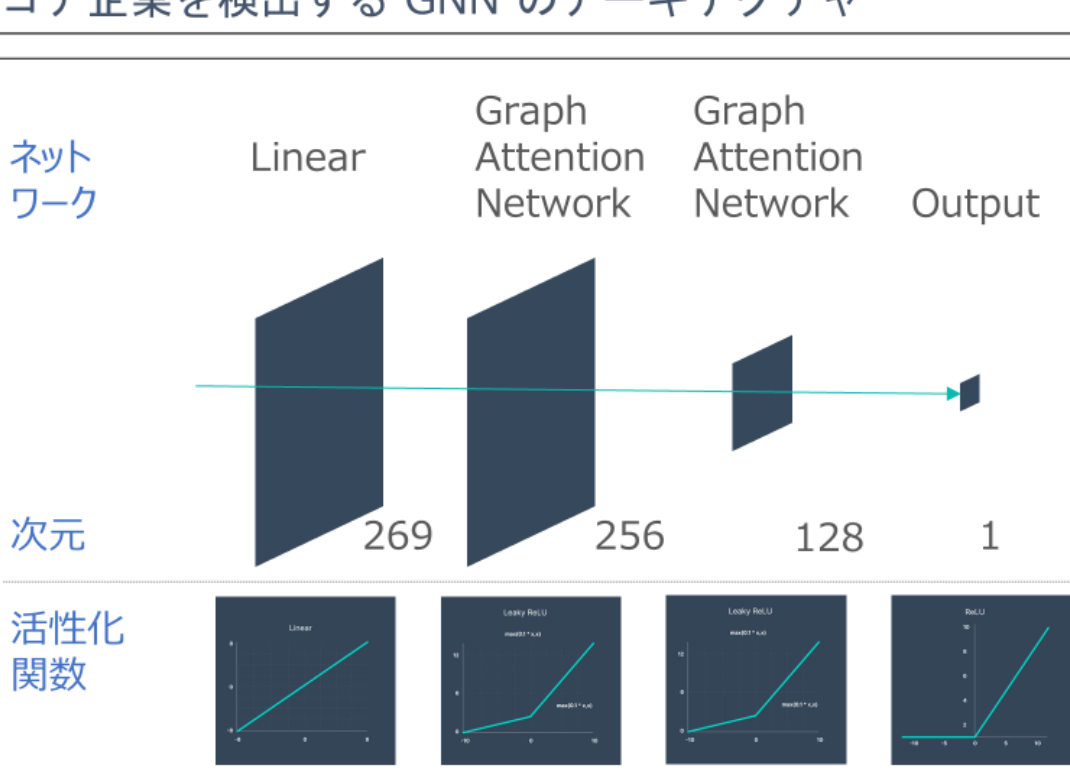
METIS は再帰的にグラフをさかのぼり、エッジ切断の数が最小となるようにグラフを分割

Luo et al. (WSDM'22)

### Graph Neural Network | コア企業の検出

線形性/非線形性を学習できるGNNのアーキテクチャを用いてスコアを学習し、コア企業を抽出

#### コア企業を検出する GNN のアーキテクチャ



活性化関数には勾配消失を防ぐため、中間層に Leaky ReLU を用いる

#### コア企業検出のためのスコアの種類

以下の値を各企業における目的変数として学習し、その予測値を重要度を表すスコアとして扱う

受発注の取引社数が多い企業を重要とみなすスコア

$$\mu_{\text{取引社数}} = \sum_{j \in I} (1[\text{受注取引}_{i,j}]) + \sum_{j \in I} (1[\text{発注取引}_{i,j}])$$

受発注の取引額が多い企業を重要とみなすスコア

$$\mu_{\text{取引額}} = \sum_{j \in I} (\text{受注取引額}_{i,j}) + \sum_{j \in I} (\text{発注取引額}_{i,j})$$

受発注の取引額/取引社数/取引コミュニティが多い企業を重要とみなすスコア

$$\begin{aligned} \mu_{\text{コア}} &= \{\sum_{j \in I} (1[\text{受注取引}_{i,j}]) + \sum_{j \in I} (1[\text{発注取引}_{i,j}])\} \\ &\quad \times \{\log(\sum_{j \in I} (\text{受注取引額}_{i,j})) + \log(\sum_{j \in I} (\text{発注取引額}_{i,j}))\} \\ &\quad \times \{\sum_{j \in I} (1[\text{異なるコミュニティ}_{i,j}])\} \end{aligned}$$

ここでは取引額の規模が重要と考え対数をとっている

## 実験結果

検出したコミュニティの地域性と、スコアが高い企業がグラフの中核になっていることを確認

#### 各コミュニティにおける企業の地理的分布



#### 各コミュニティの地域解釈

北海道	東京 大坂 愛知 九州	東京 福井 愛知 大坂	東京 大坂
東京 大坂 愛知 東京	東京 大坂 大坂 東京	山梨 東京	東京 神奈川 大坂 福岡
大坂 東京 愛知 広島 福岡	神奈川 東京	富山 福井 石川 千葉 京都	福岡 熊本 長崎 大分
長野 静岡 愛知 岐阜	兵庫 大坂 岡山 四国	福島 新潟 宮城 (東北)	東京 東京
東京 大坂 東京 大坂 大坂 東京	東京 大坂 東京 大坂 東京	茨城	愛知 東京

#### 北海道コミュニティに含まれる北海道の企業間取引ネットワーク

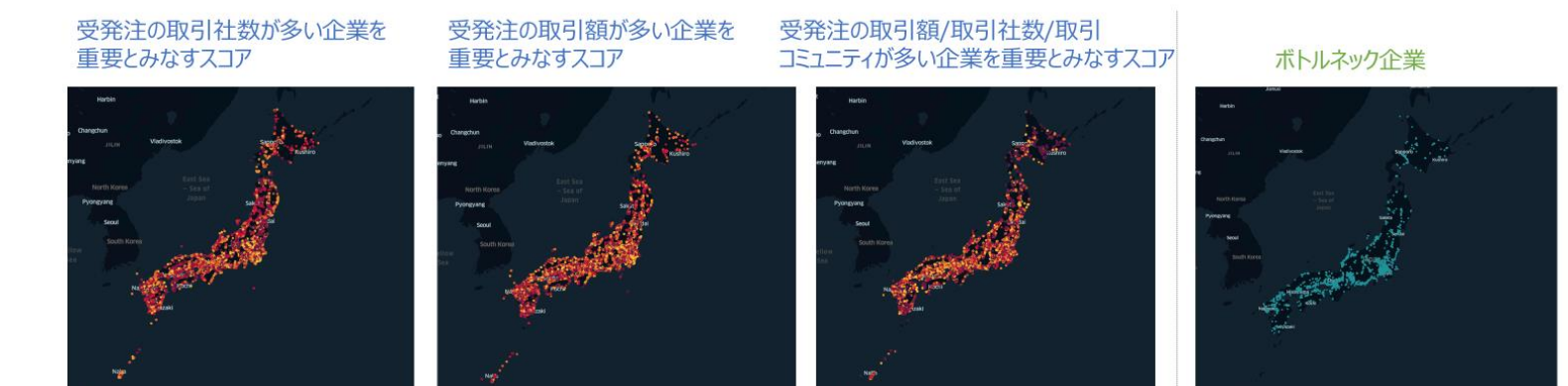


#### 東京コミュニティに含まれる北海道の企業間取引ネットワーク



#### 実験と考察 | コア企業の検出

各スコアが上位95%である企業を抽出。コミュニティを考慮したスコアが高い企業は多くの企業と取引していることを確認



スコア	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数
'01%'	2	25.6	2	2	20.1	2	4	27.1	3	0	0.0	0
'05%'	5	87.0	3	5	75.0	3	7	81.9	4	0	0.0	0
'25%'	14	432.6	7	13	405.6	6	17	408.1	8	5	185.5	3
'50%'	24	1481.7	10	24	1544.8	10	28	1460.5	11	22	1584.2	8
'75%'	45	6926.4	15	48	7548.6	16	51	6464.1	16	51	9514.4	16
'95%'	225	89456.3	32	239	94741.9	33	228	76221.1	33	288	117295.2	36
'99%'	828	468066.4	47	912	555284.3	48	853	488163.8	47	1035	708265.4	49