



大規模SNS研究でのグラフ活用

だけじゃない

奈良先端科学技術大学院大学
ユビキタスコンピューティングシステム研究室
助教 諏訪博彦



- 石原裕規, 諏訪博彦, 鳥海不二夫, 太田敏澄, 東日本大震災前後における重要アカウントの抽出とコミュニケーション形態の変容, 電子情報通信学会論文誌D, Vol.J99-D, No.5, pp.501-513, 2016.
- 鳥海不二夫, 山本仁志, 諏訪博彦, 岡田勇, 和泉潔, 橋本康弘, 大量SNSサイトの比較分析, 人工知能学会論文誌, 25巻, 1号, pp.78-89, 2010.
- 山本仁志, 諏訪博彦, 岡田勇, 鳥海不二夫, 和泉潔, 橋本康弘, コミュニケーション構造の推移による大量SNSの分類, 日本社会情報学会誌, 23巻1号, pp33-43, 2011.
- 佐藤智行, 小川祐樹, 諏訪博彦, 太田敏澄, アーティストネットワークを用いたインディーズアーティスト推薦システムの構築, 情報処理学会論文誌, Vol.52, No.1, pp.44-55, 2011.
- 三浦大樹, 諏訪博彦, 鳥海不二夫, 鬼塚真, ソーシャルサーチのための効率的な検索アルゴリズムの提案, 情報処理学会誌 データベース, Vol.6, No.3, pp.23-29, 2013.
- Kotone Senju, Teruhiro Mizumoto, Hirohiko Suwa, Yutaka Arakawa, Keiichi Yasumoto, Designing Strategy for Resolving Maldistribution of Vehicles in One-Way Car-Sharing through Active Trip Request to Potential Users, The workshop on Human-Centered Computational Sensing (HCCS'18)



グラフマイニングで何ができるか？

理解する

活用する



震災時の情報流通を支える Twitterアカウントの発見

石原裕規^{*1} 諏訪博彦^{*1} 鳥海不二夫^{*2} 太田敏澄^{*1}

^{*1}電気通信大学 ^{*2}東京大学



東日本大震災時に、Twitterは情報流通ツールとして活用[総務省]
その情報流通において

- 重要だったアカウントはどんなアカウントか？
- どんなコミュニケーションをとっていたのか？

今後の災害時における
Twitterの有効活用に活かせる



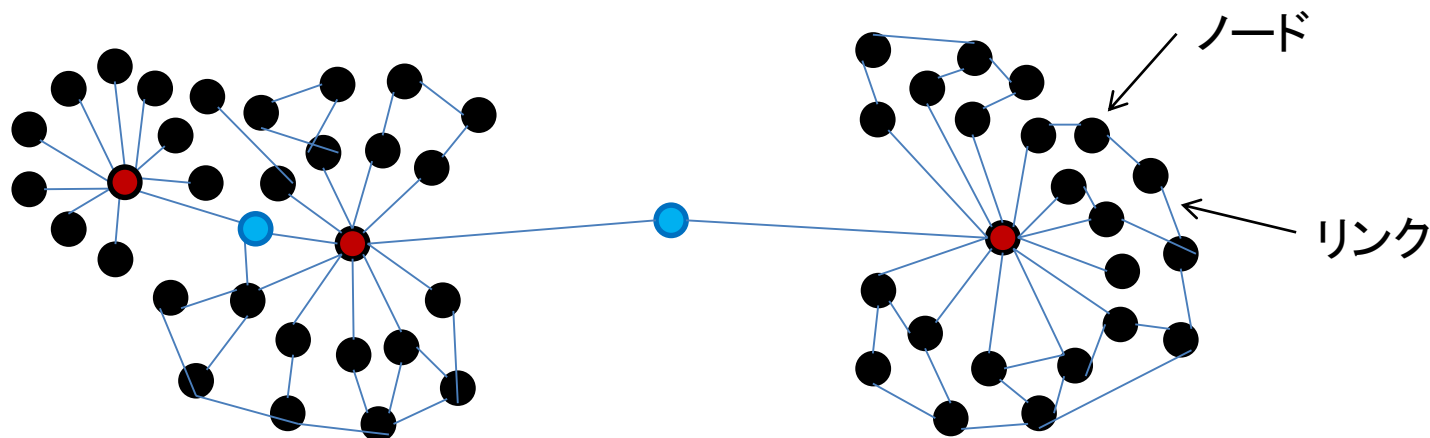
帰宅困難[47news]



震災に関する情報は人々のコミュニケーションにより流通している

- 情報拡散の起点となるアカウント(ハブ)
- 情報を仲介(繋げる)するアカウント

これらのアカウントを特定することは情報流通の点で重要





- 重要なアカウントを発見する
 - 情報拡散の起点となるアカウント
 - 情報を仲介するアカウント
- やったこと
 - 震災前後で抽出される重要アカウントの比較
 - 重要アカウントのコミュニケーション形態を明らかにする



▶ ネットワーク分析を用いて中心的なアカウントの特定する

- **次数中心性(Degree Centrality)**

- 情報拡散の役割の強さを示す
- 各ノードの持つ関係の多寡により, 重要性を評価する指標

$$DC(v) = \frac{degree(v)}{N-1}$$

- **媒介中心性(Betweenness Centrality)**

- 情報と情報を繋ぐ役割の強さを示す
- 他のノードとノードを結ぶ経路上にあるノードを高く評価する指標
- 全対全の最短経路の計算が必要

近似解(SSCA#2)

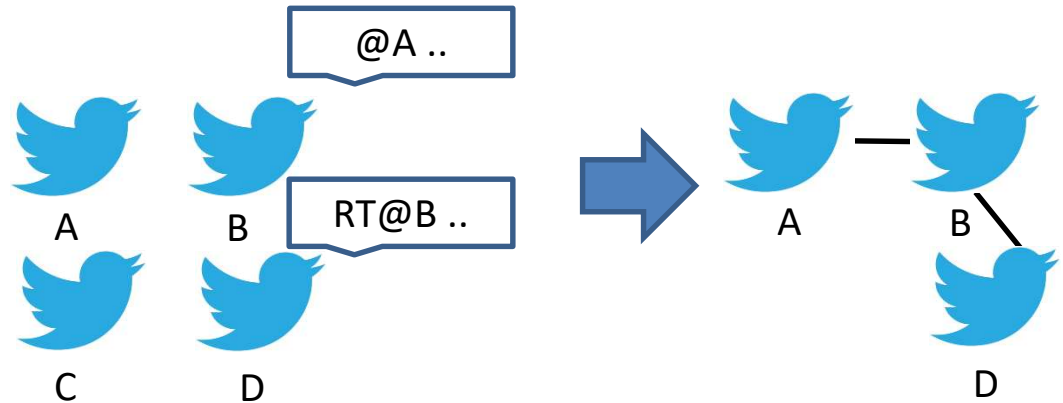
$$BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

ネットワーク生成



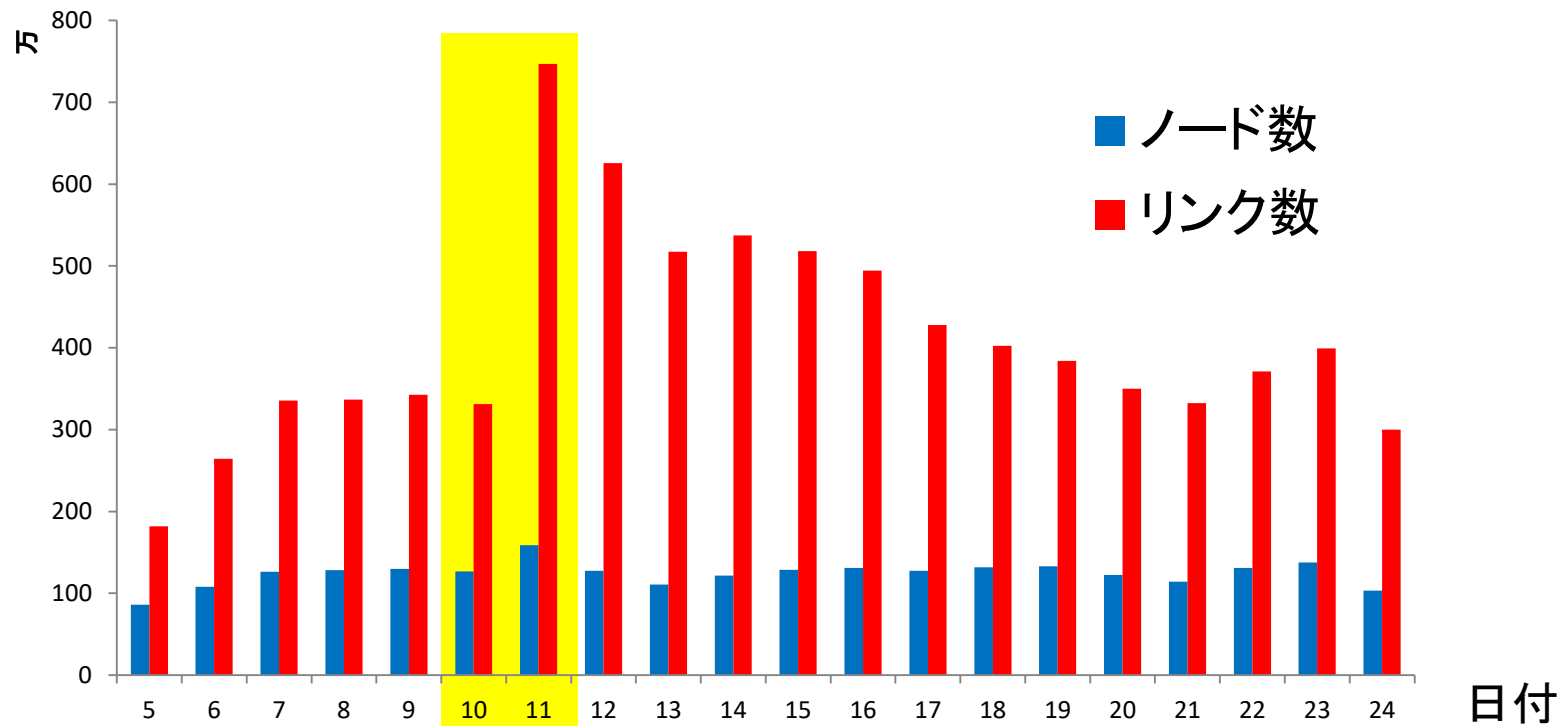
- データ

- 収集期間
 - 2011年3月5日から3月24日(20日間)
- 収集条件
 - 日本語で、ある時点においてツイート数200以上のアカウント
- 収集ツイート数
 - 3億3235万199件



- 一日毎のコミュニケーションネットワークの生成
 - @アカウント名をもとに生成
 - 無向グラフ
- ※フォロー, フォロワーネットワークではない
- 一日平均, 約124万アカウントが繋がっている

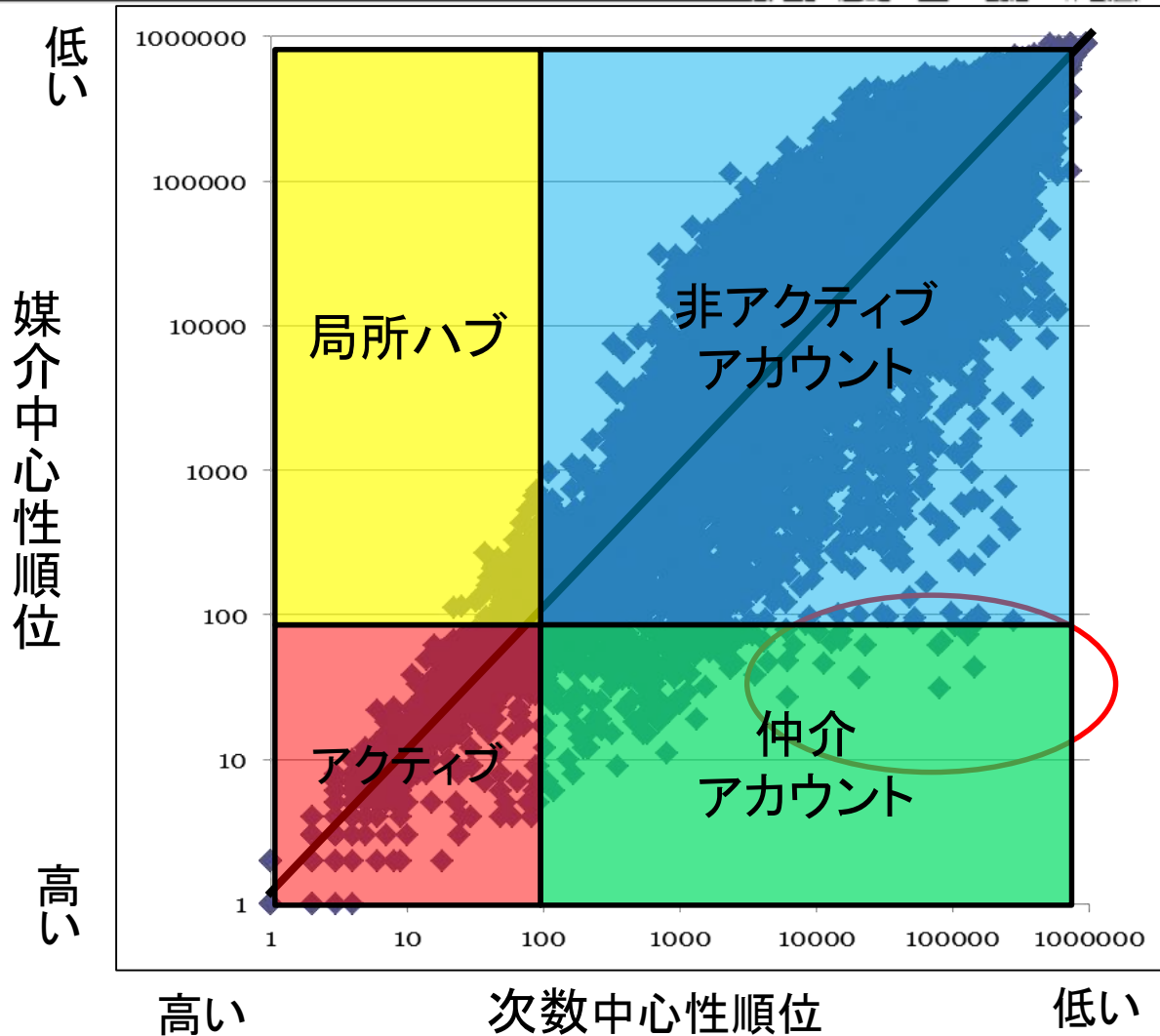
ノード数,リンク数の推移



震災前後におけるノード数とリンク数の変化

震災後(11日)にリンク数が増加したことがわかる
このことから、コミュニケーションが活発化したことがわかる。

次数中心性と媒介中心性



・高い相関を示している

震災前後におけるアクティブアカウントの一例

分類	震災前	震災後
ニュース 新聞社 アカウント	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社)	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社:ニュースサイト) nhk_seikatsu(NHK生活情報部) nhk_news(NHKニュース) nhk_kabun(NHK科学文化部) nhk_HORIJUN(NHK堀潤) Asahi_Shakai(朝日新聞社会部)
個人 アカウント	masason(孫正義:実業家) tsuda(津田大介) ayu_19980408(浜崎あゆみ)	masason(孫正義:実業家) tsuda(津田大介:ジャーナリスト) ayu_19980408(浜崎あゆみ:歌手) hayano(早野龍五:大学教授)..
その他	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) shuumai(人工知能bot) wwwwww_bot(ネタbot)	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) touhokujishin(震災の情報センタ) OfficialTEPCO(東京電力) Kantei_Saigai(首相官邸災害) team_nakagawa(東大放射能治療チーム) radio_rfc_japan(ラジオ福島)

震災前後でアクティブ

震災前後におけるアクティブアカウントの一例

分類	震災前	震災後
ニュース 新聞社 アカウント	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社)	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社:ニュースサイト) nhk_seikatsu(NHK生活情報部) nhk_news(NHKニュース) nhk_kabun(NHK科学文化部) nhk_HORIJUN(NHK堀潤) Asahi_Shakai(朝日新聞社会部)
個人 アカウント	masason(孫正義:実業家) tsuda(津田大介) ayu_19980408(浜崎あゆみ)	masason(孫正義:実業家) tsuda(津田大介:ジャーナリスト) ayu_19980408(浜崎あゆみ:歌手) hayano(早野龍五:大学教授)
その他	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) shuumai(人工知能bot) wwwwww_bot(ネタbot)	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) touhokujishin(震災の情報センタ) OfficialTEPCO(東京電力) Kantei_Saigai(首相官邸災害) team_nakagawa(東大放射能治療チーム) radio_rfc_japan(ラジオ福島)

震災後アクティブ

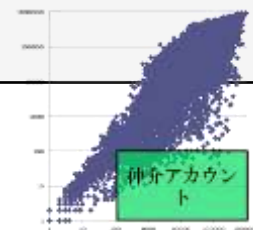
震災前後におけるアクティブアカウントの一例

分類	震災前	震災後
ニュース 新聞社 アカウント	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社)	NHK_PR(NHK_広報局) 47news(共同通信ニュース) asahi(朝日新聞社:ニュースサイト) nhk_seikatsu(NHK生活情報部) nhk_news(NHKニュース) nhk_kabun(NHK科学文化部) nhk_HORIJUN(NHK堀潤) Asahi_Shakai(朝日新聞社会部)
個人 アカウント	masason(孫正義:実業家) tsuda(津田大介) ayu_19980408(浜崎あゆみ)	masason(孫正義:実業家) tsuda(津田大介:ジャーナリスト) ayu_19980408(浜崎あゆみ:歌手) hayano(早野龍五:大学教授)..
その他	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) shuumai(人工知能bot) wwwwww_bot(ネタbot) ボットによる違い	Kopipedoujou(コピペ道場bot) sazae_f(サザエさんbot) touhokujishin(震災の情報センタ) OfficialTEPCO(東京電力) Kantei_Saigai(首相官邸災害) team_nakagawa(東大放射能治療チーム) radio_rfc_japan(ラジオ福島)

仲介アカウント



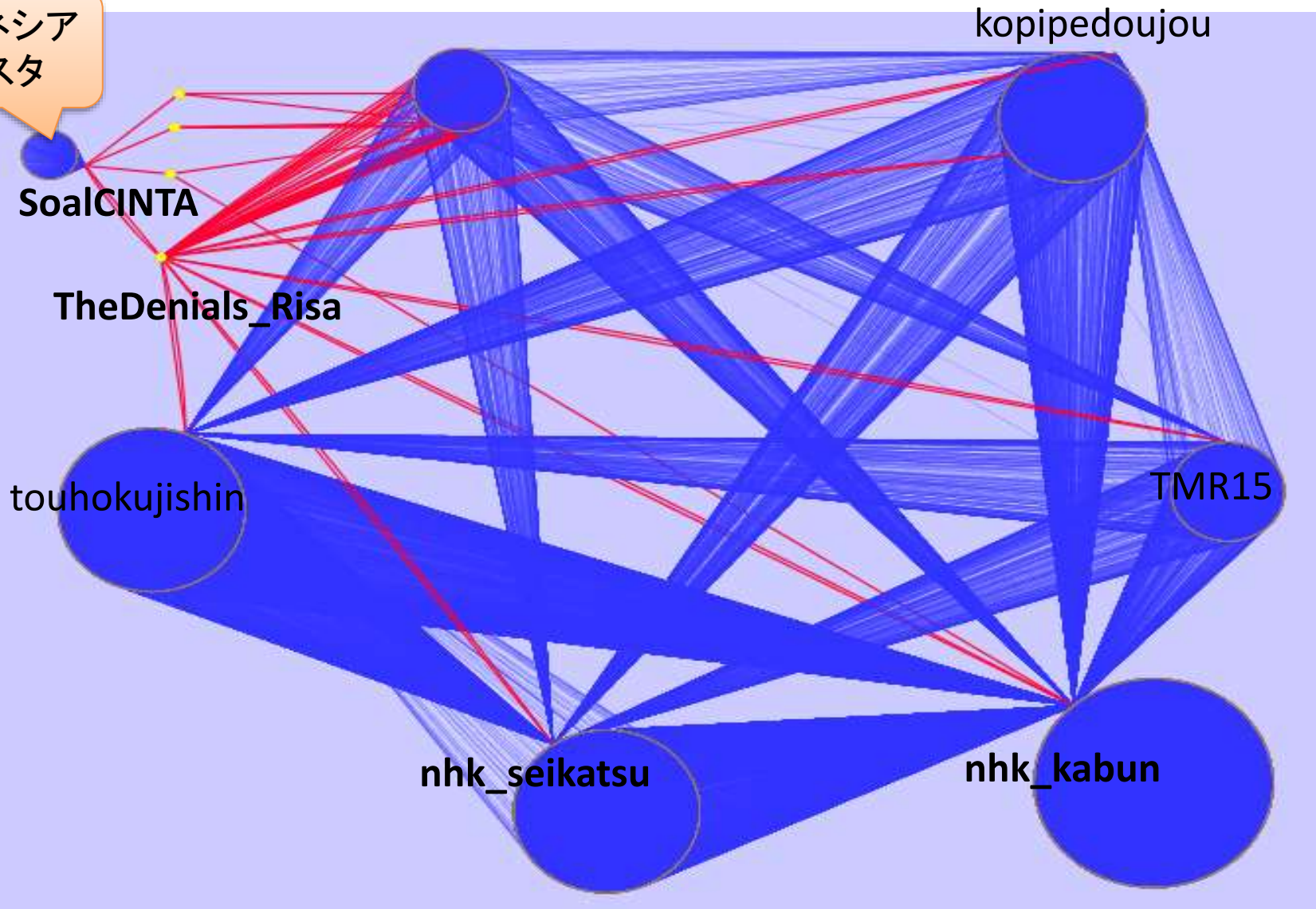
分類	震災前	震災後
個人 アカウント	ladygaga(レディガガ:歌手) detikcom(インドネシアのサービス) 23kin (複数の言語でツイート) Charliesheen(チャーリー・シーン:俳優)	ladygaga(レディガガ:歌手) mjjeje(韓国人歌手) detikcom(インドネシアのサービス) TheDenials_Risa
その他	viratter (RTされてるツイートをRT) retweeterjp (RTされてるツイートをRT)	該当なし



- 度数中心性が比較的高いアカウントを仲介しているアカウント(viratter, retweeterjp)が抽出
- 仲介アカウントは海外アカウントが多い

仲介アカウントのネットワーク例

インドネシア
クラスター





ネットワーク構造分析だけでよいか？

重要アカウントがわかればよいか？

重要アカウントはどんなコミュニケーション？

分析目的

震災前後で重要アカウントはどのようなコミュニケーションにより，中心性を高めているのか明らかにする

方法

コミュニケーション形態をもとに，アカウントを分類し，震災前後で比較する．

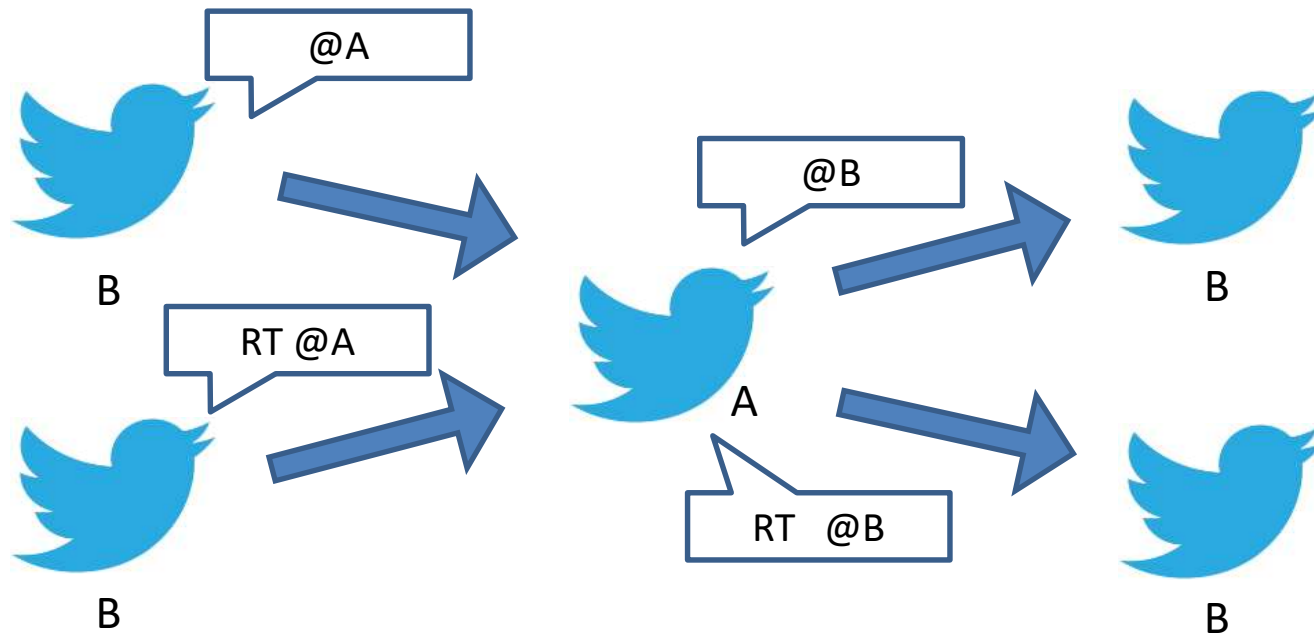
抽出条件:

次数中心性または媒介中心性が100位以上のアカウント

震災前:2011年3月6日～10日 315アカウント

震災後:2011年3月11日～15日 396アカウント

分類のための指標



$\text{Out_Reply率} = \text{Out_Reply数} / \text{総コミュニケーション数}$

$\text{Out_RT率} = \text{Out_RT数} / \text{総コミュニケーション数}$

$\text{In_Reply率} = \text{In_Reply数} / \text{総コミュニケーション数}$

$\text{In_RT率} = \text{In_RT数} / \text{総コミュニケーション数}$

$\text{総コミュニケーション数} = \text{Out_Reply数} + \text{Out_RT数} + \text{In_Reply数} + \text{In_RT数}$

コミュニケーション形態の分析(k-means)

	クラスタ(K=4, 母数=711 前:315 後:396)			
	Out_Reply型	Out_RT型	In_Reply型	In_RT型
Out_reply率	0.74	0.06	0.04	0.02
Out_RT率	0.06	0.78	0.02	0.01
In_Reply率	0.13	0.03	0.62	
In_RT率	0.06	0.13	0.31	
該当アカウント数 (%)	29 (4%)	26 (4%)	77 (11%)	
震災前 (%)	29 (9%)	16(5%)	67(21%)	203(64%)
震災後 (%)	0 (0%)	10		376(96%)
前アカウント	FlandreBot ReiuziUtsuhoBot	habu_b nannsuba_bot	R15 ayu_980408	sazae_f asahi NHK_PR
後アカウント	該当無	Litha_Sacra cuppekezuke	ladygaga justinbieber mjjeje	Kantei_Saigai hayano asahi NHK_PR

In_RT型が増加

媒介中心性が高い

考察(1/2)



- 震災前は数多くのBotが抽出されたが，震災後は，殆ど抽出されなかった.
- 自粛ムードのため，管理者がツイートを停止
- 震災後は，震災に関する情報を発信するアカウントが重要



楽しむためのツールから情報収集ツールへ

- 震災後に抽出された重要アカウントとして，NHK，hayano，元botや浜崎あゆみなどが抽出された



震災後では，信頼性，専門性の高いアカウントが重要
災害前に発信力のあるアカウントが重要であることがわかった



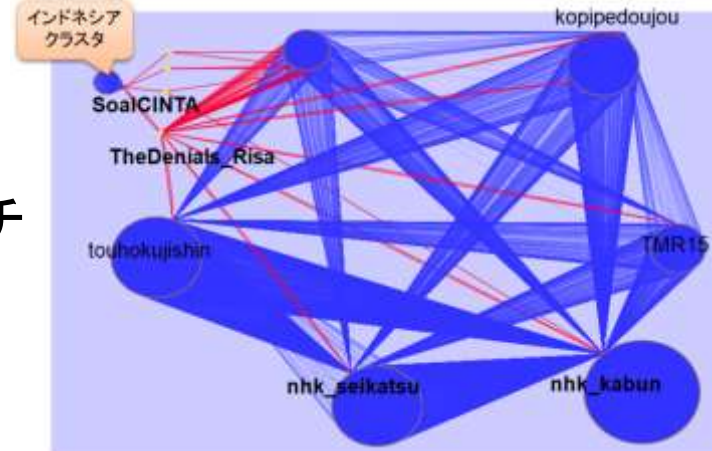
- コミュニケーション形態に基づく分析より、震災後における重要アカウントは、RTにより、中心性が高くなっていることが確認されている



震災後はリツイートが重要な情報流通手段となっている



- 緊急/重要情報の拡散支援
 - アクティブアカウントに対するアプローチ
 - 通常時に発信力のあるアカウント（人気botや有名人）による緊急情報の拡散
 - 信頼性の高いアカウントによる重要情報の拡散
- ▶ マイノリティグループに対する支援
 - 仲介アカウントに対するアプローチ
 - TheDenials_Risaにアプローチ
 - インドネシアクラスタへアプローチすることができる



グラフマイニングは何に役に立ったか？



知見1

- 震災時のTwitter上のコミュニケーションを理解する
- 重要アカウントを見つける

知見2

- グラフマイニングだけでは、コミュニケーションは理解できない



- 他の分析手法との組み合わせが大切

(ここではK-means)

理解する



グラフマイニングで何ができるか？

理解する

活用する

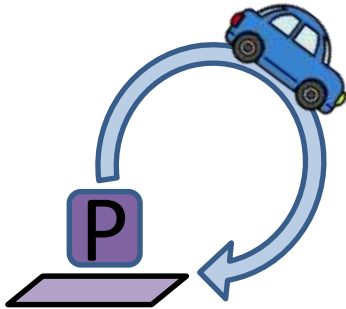


グラフマイニングで何ができるか？

理解する

One-Way カーシェアリングにおける利用者分析

[ラウンドトリップ式]

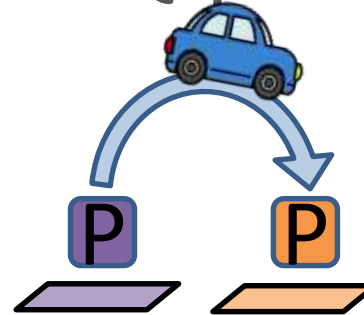


- 貸出場所 = 返却場所
- 現在主流の方式

返却場所へ戻る
時間や手間が面倒



[ワンウェイ方式]



乗り捨て
可能

- 貸出場所 ≠ 返却場所
- 2014年から
サービス運営可能

利用者の
行動の幅が広がる



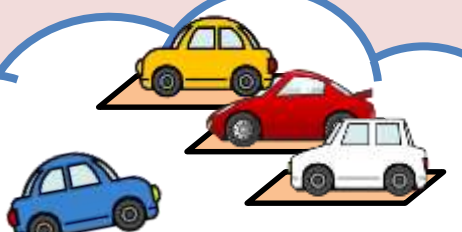
ワンウェイ方式の問題点



乗りたい場所に
車がない



利用者

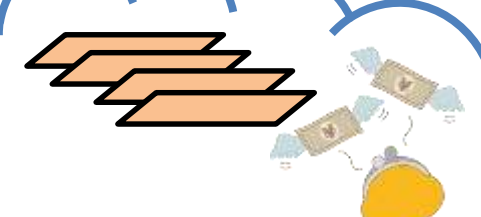


乗り捨てスペースがない

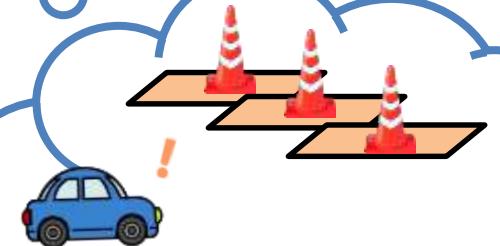
車両の偏りが問題
→予約受諾率の低下



運営会社



専用スペース
確保のコスト負担



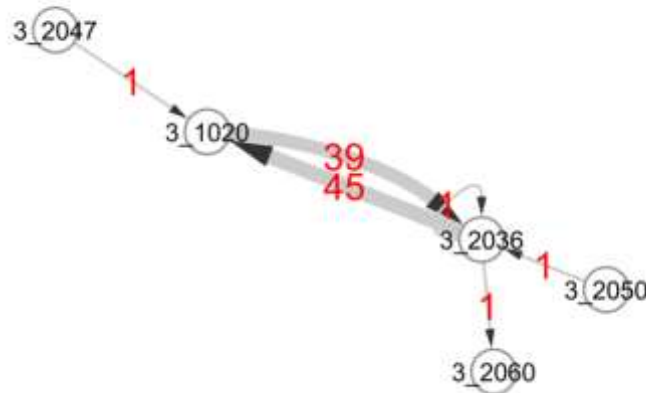
一般車の駐車不可

利用者に移動を依頼. 誰に頼む?

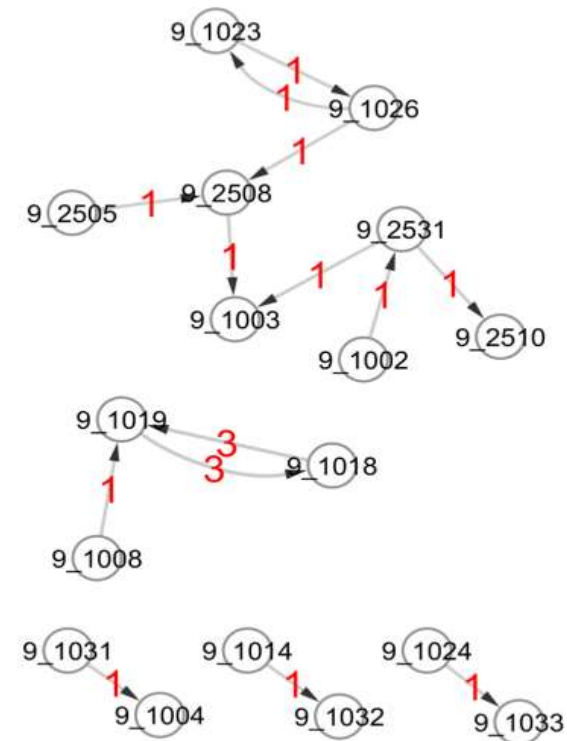
利用者タイプの可視化



	平均 (130件)	クラスタ A (19件)	クラスタB (15件)	クラスタC (27件)	クラスタ D (43件)	クラスタE (26件)
NetRange	2.46	2.26	1.73	4.14	2.39	1.42
Cnt_station	7.38	7.10	5.13	13.2	6.88	3.65
Log_cnt_rt	2.64	3.69	2.09	2.96	2.16	2.66
Per_rt_round	0.02	0.02	0.01	0.02	0.02	0.03
Per_two_way	0.17	0.30	0.46	0.12	0.04	0.18
Per_rt0	0.35	0.36	0.33	0.18	0.27	0.65
Per_rt1	0.20	0.27	0.22	0.12	0.18	0.24
Per_rt2	0.10	0.11	0.15	0.08	0.14	0.04



クラスタA: 固定

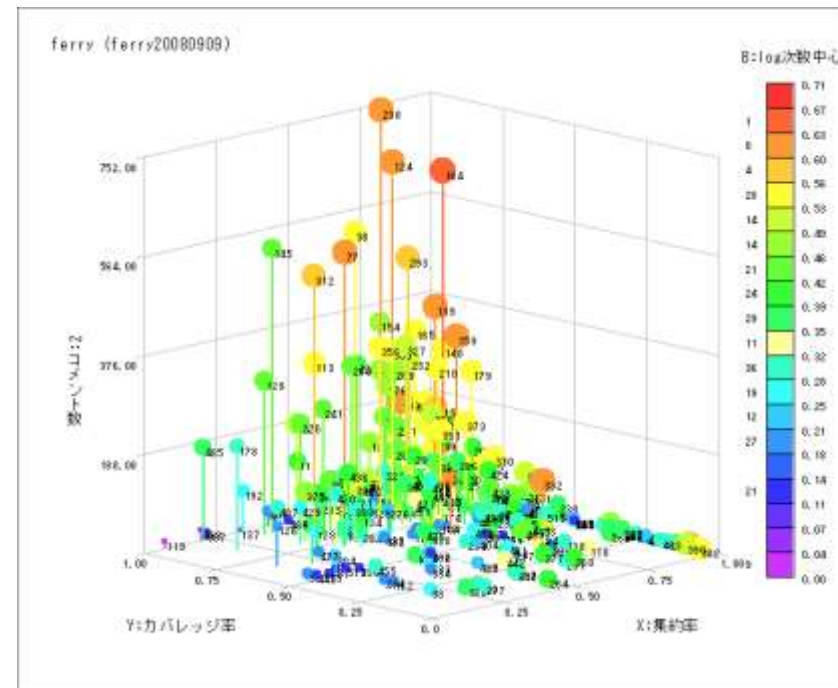


クラスタC: 分散型

大量小規模SNSの分析



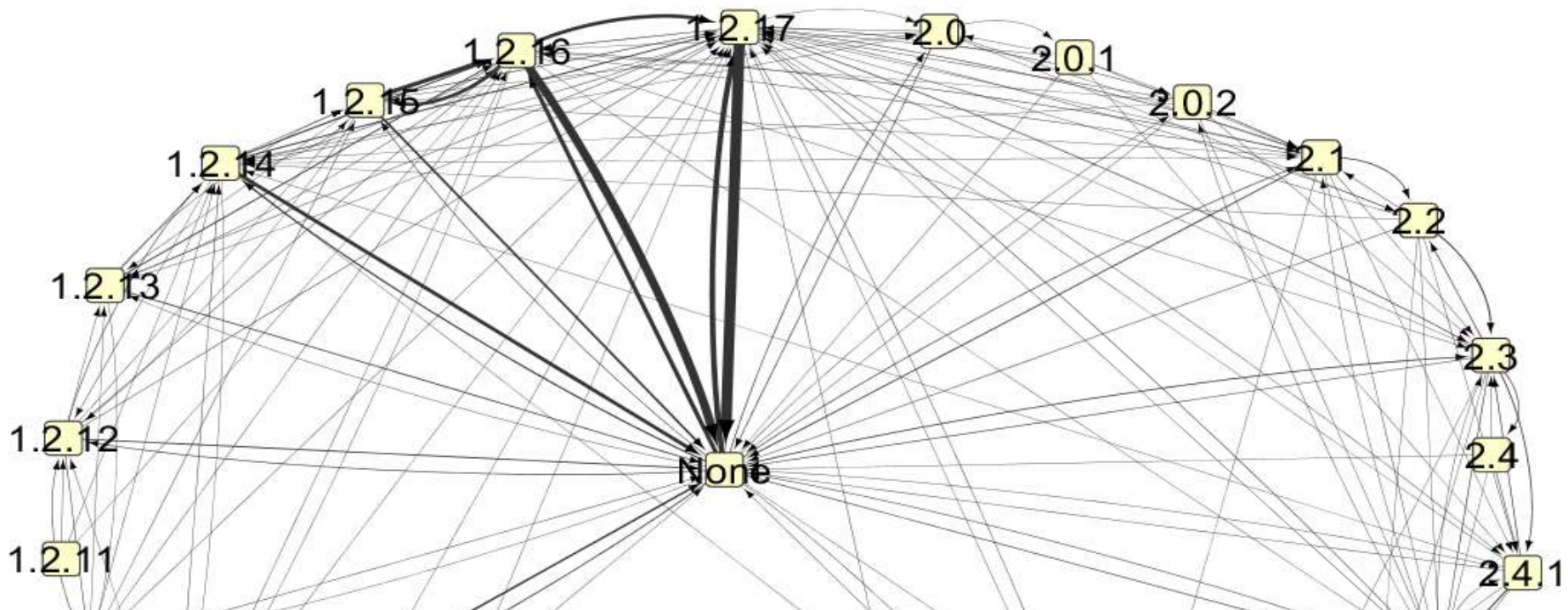
- ソネットエンタテインメントと5大学および産総研が連携した共同研究
- So-net SNSのデータベースに蓄積されたデータを対象
 - 投稿, 友人関係, コメントなどを利用
- ネットワーク, コミュニケーション, 成長率の観点から分析
 - 大規模SNS (mixi) とは構造が異なる
 - 開設時の幅広いコミュニケーションが成長に寄与
 - 現実の人間関係がベースとなるSNSは限定的なコミュニケーションにより維持



山本仁志, 諏訪博彦, 岡田勇, 鳥海不二夫, 和泉潔, 橋本康弘, コミュニケーション構造の推移による大量SNSの分類, 日本社会情報学会誌, 23巻1号, pp33-43, 2011.

鳥海不二夫, 山本仁志, 諏訪博彦, 岡田勇, 和泉潔, 橋本康弘, 大量SNSサイトの比較分析, 人工知能学会論文誌, 25巻, 1号, pp.78-89, 2010.

OSSにおけるライブラリ進化に関する分析



log4jは, 1.2.17が安定板
2.X系から1.2.17に戻ることも多数

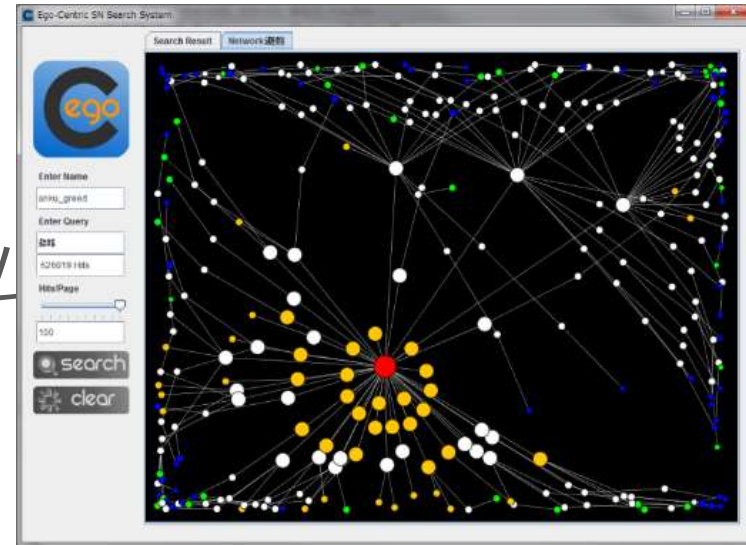


グラフマイニングで何ができるか？

活用する



- ソーシャルサーチの高速化
 - ソーシャルサーチ：人の繋がりを考慮した検索
- 3つのアルゴリズムを提案
 - Single Index アルゴリズム
 - Social Index Graph アルゴリズム
 - ハイブリッドアルゴリズム
- Twitterデータを用いて評価
 - ユーザ数約10万，投稿数約7000万
 - ヒット数の多寡による性能特性を確認

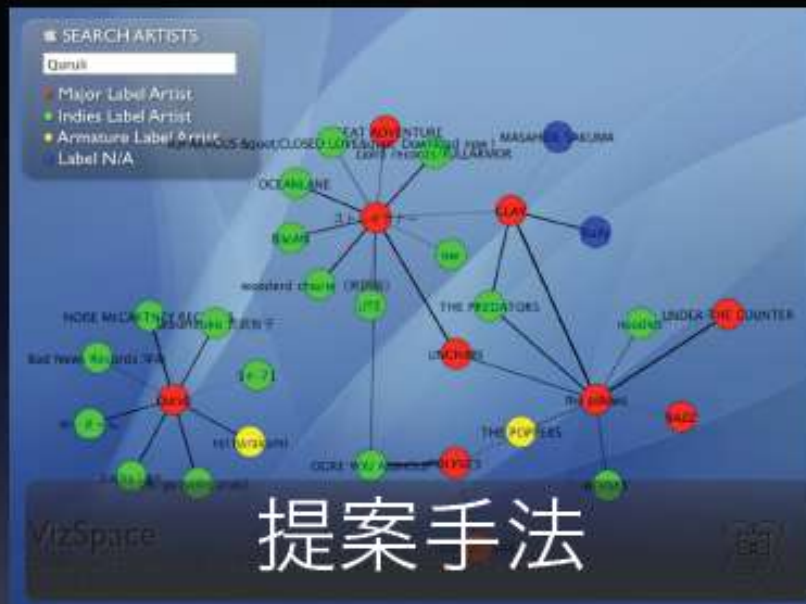


データベースにおけるIndexの作り方を工夫

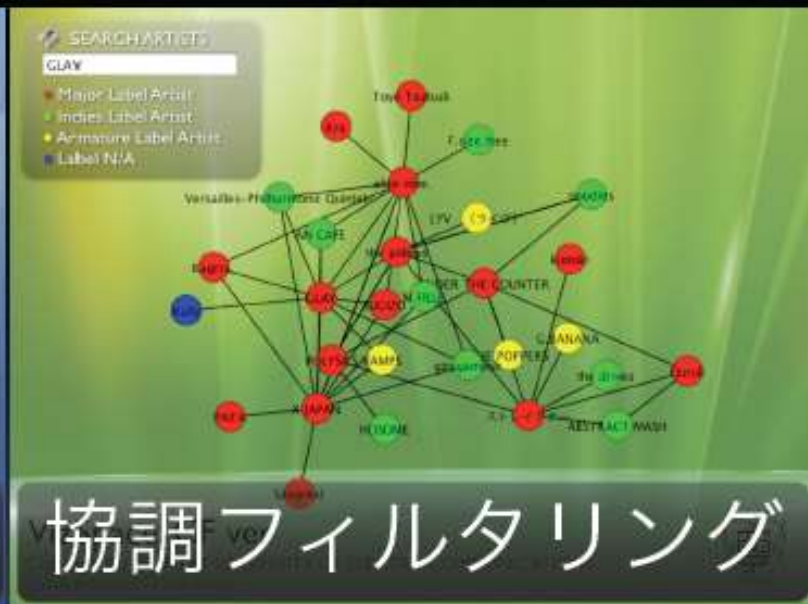
三浦大樹, 諏訪博彦, 鳥海不二夫, 鬼塚真, ソーシャルサーチのための効率的な検索アルゴリズムの提案, 情報処理学会誌 データベース, Vol.6, No.3, pp.23-29, 2013.

ネットワーク構造を用いたインディーズ推薦

同一シードからの推薦例(the pillows, Quruli, GLAY, ストレイテナー)



メジャー10、インディーズ18
少数のノード毎に島を形成している



メジャー15、インディーズ9
少ないノードでリンクを張り合ってしまう

また、推薦結果の一致率は10%未満
→さらに推薦特性について分析を行う

- メジャー
- インディーズ
- アマチュア
- 分類無し

道路交通網のボトルネックを考慮した避難誘導

東日本大震災における帰宅困難者数: 400万人¹⁾

自力の帰宅が困難となり、保護が必要になる人々



収容数超過によるたらい回し発生



被災者の一斉避難による混雑の発生

本研究で対象とする問題

大都市や観光地を対象にした避難誘導

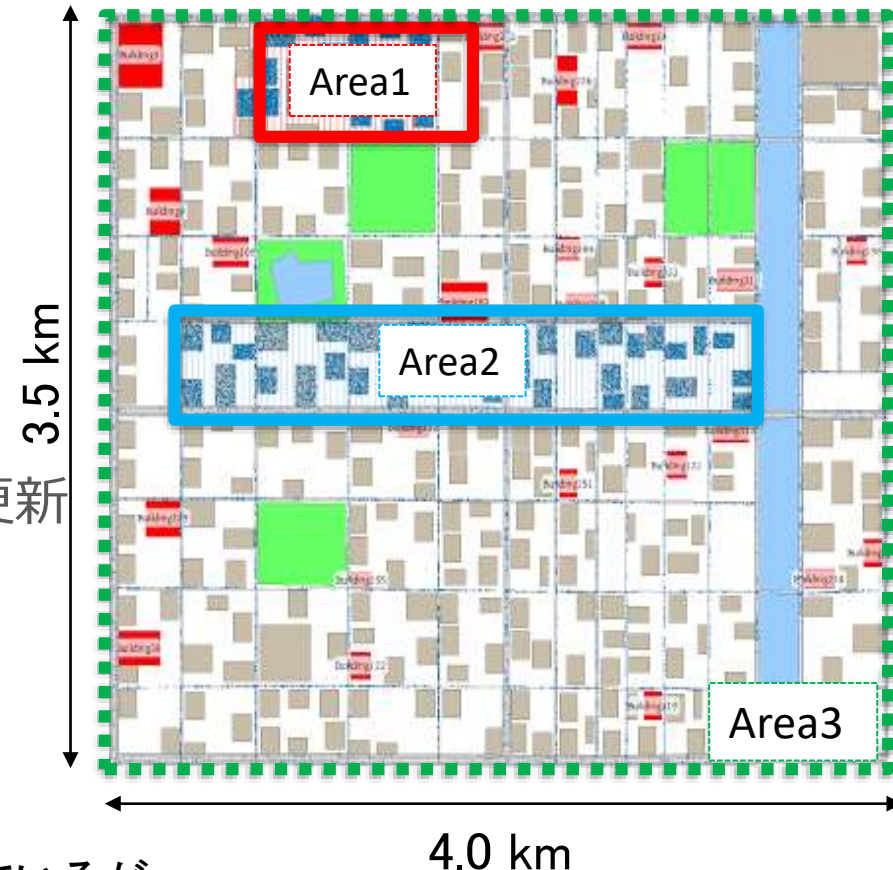
※人が一部地域に密集している場所

¹⁾ 廣井悠ら: 東日本大震災における首都圏の帰宅困難者に関する社会調査, 地域安全学会論文集, Vol.15, pp. 343-354, 2011

道路交通網のボトルネックを考慮した避難誘導

- 道路交通網上での避難シミュレーション

- 京都市，祇園祭を想定
- 被災者：6万人
- Area1,Area2に人が集中
- 地震発生時に避難場所を推薦
- 提案手法
 - 初期配置からシミュレーション開始
 - ボトルネックを考慮した避難時間に更新
 - 繰り返しシミュレーション
 - 避難時間が短くなる避難所を推薦
- 課題：計算に時間がかかる



現在は，グラフを用いずにシミュレーションしているが，
グラフマイニング技術は高速化に貢献



グラフマイニングで何ができるか？

理解する

活用する



- (何かを) 理解する
 - ネットワーク自体
 - 指標化（次数中心性など，スケールフリー性など），可視化など
 - ネットワーク構造をもつ何か
 - SNS，人間関係，企業提携，道路交通網，神経ネットワークなど
- (何かを) 予測する
 - ネットワークの成長/衰退予測
 - SNSの成長予測
 - 友人関係を予測
 - ネットワーク構造上の何かを予測
 - 友人関係から嗜好予測
 - SNS上での情報拡散を予測
 - 道路交通網上の渋滞予測
- (何かを) 推薦する
 - ネットワークの成長支援
 - 効果的なSNSの運用法の推薦
 - 友達推薦
 - ネットワーク構造に基づいた推薦
 - アーティストの推薦
 - 効果的な情報拡散手法の推薦
 - 効果的な避難経路の推薦



将来の活用構想

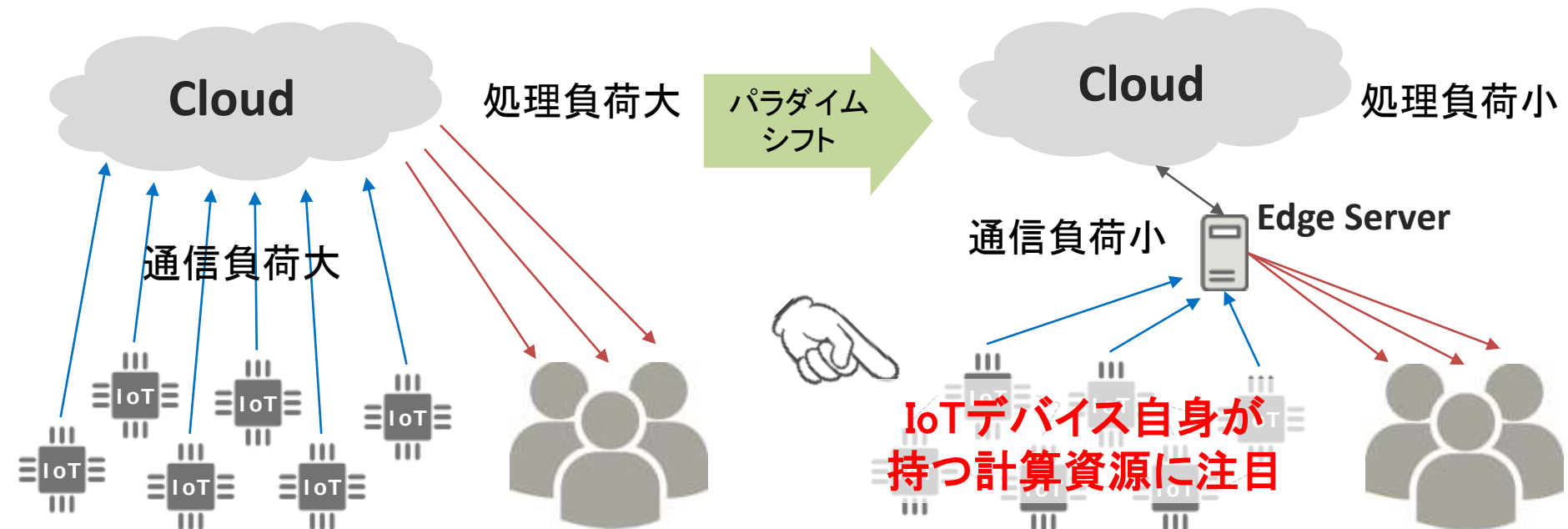
IFoT: IoTデータ流の地産地処を実現する
IoTデバイス向け分散処理プラットフォーム



奈良先端科学技術大学院大学
ユビキタスコンピューティングシステム研究室
安本慶一

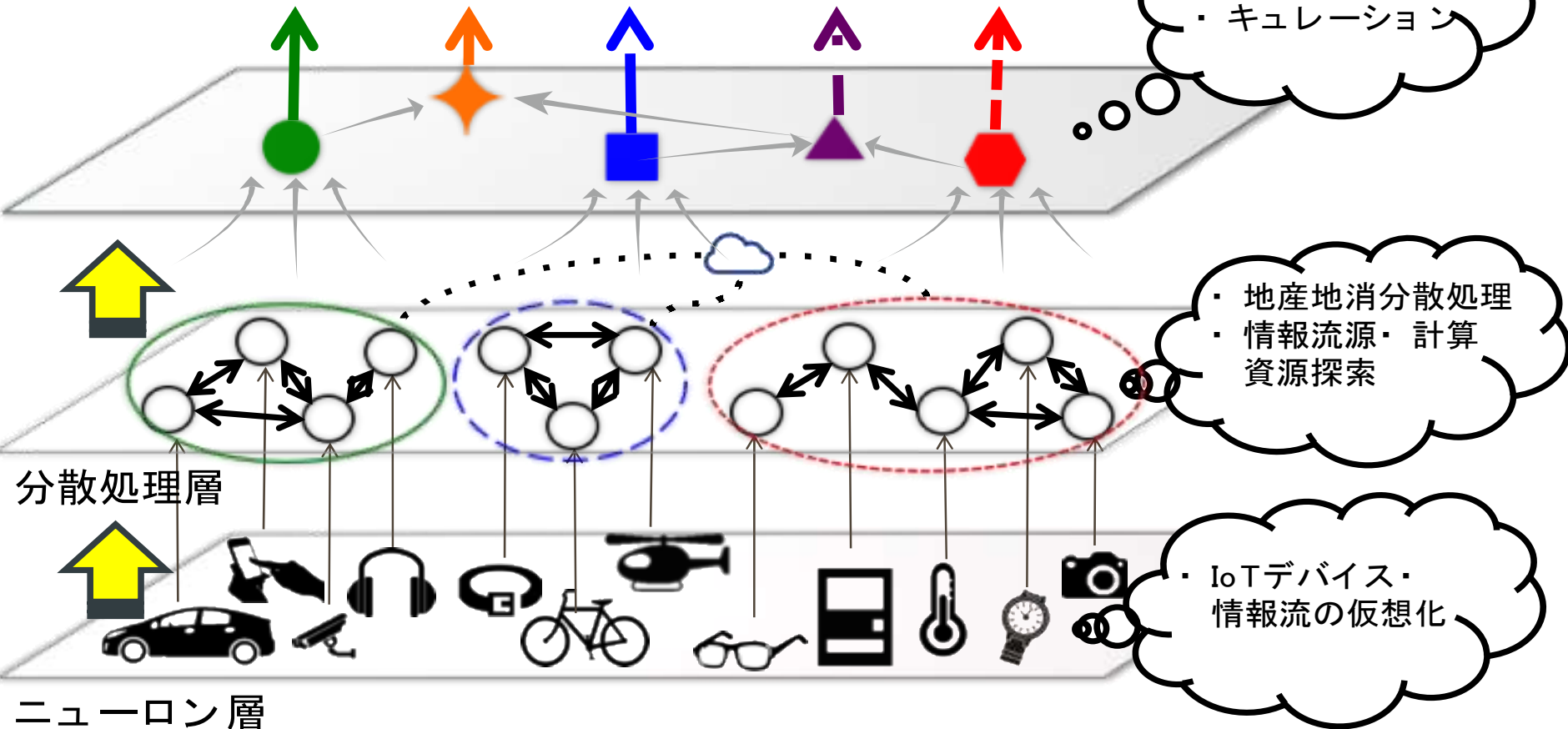


- 2020年：社会に普及するIoTデバイス 500億台
- 情報流・エッジコンピューティングの概念が注目
→要求：IoTデータ流を可能な限りエッジ側で処理



IFoT: Information Flow of Things

知的処理層

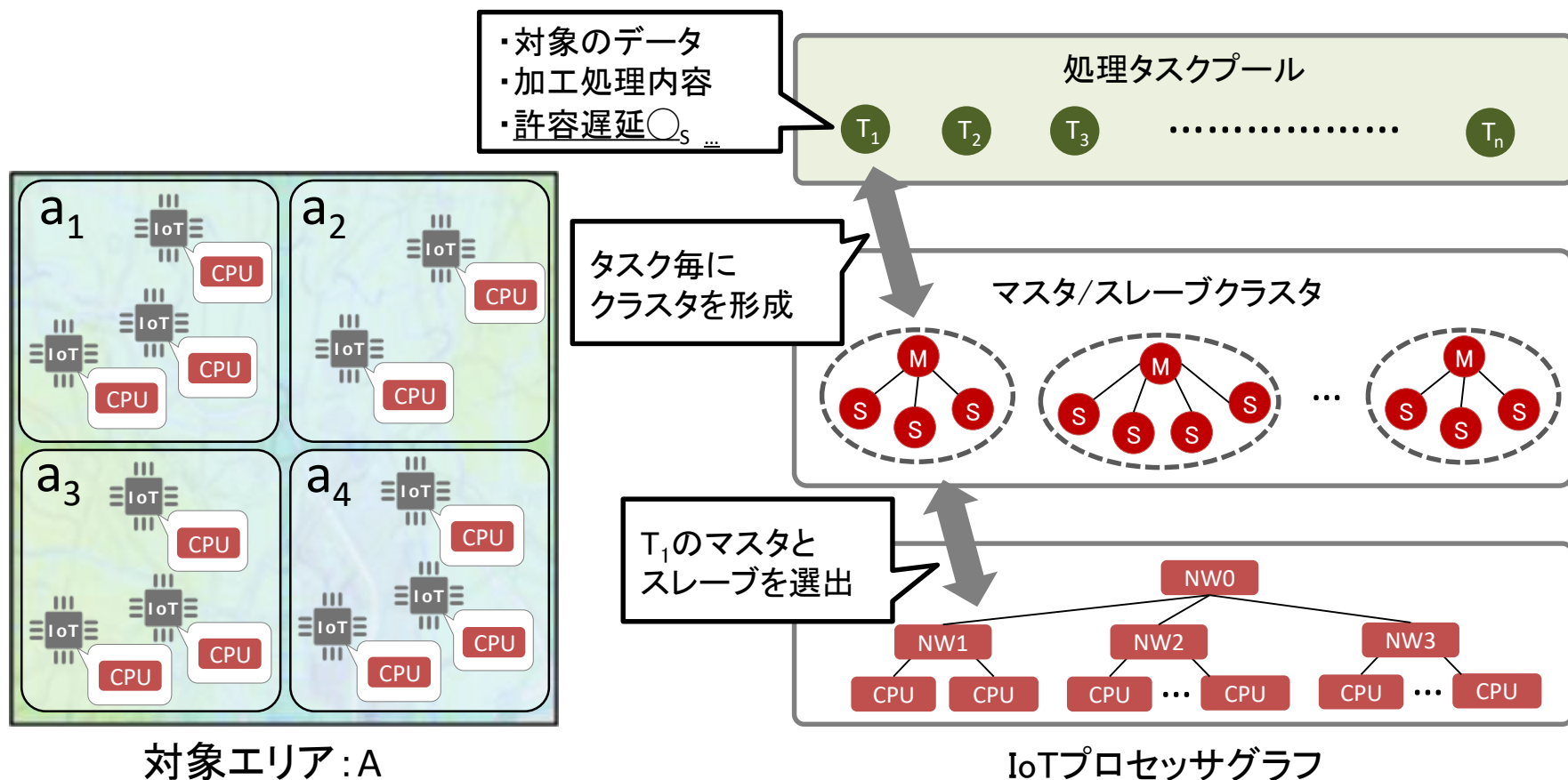


Keiichi Yasumoto, Hirozumi Yamaguchi, Hiroshi Shigeno:
[Invited Paper] [Survey of Real-time Processing Technologies of IoT Data Streams](#),
Journal of Information Processing, Vol.24, No.2, pp.195-202 (2016).

技術的課題1: IoTデバイス群の計算資源をどう集約するか？

➤ マスタ/スレーブクラスタの動的形成による分散処理

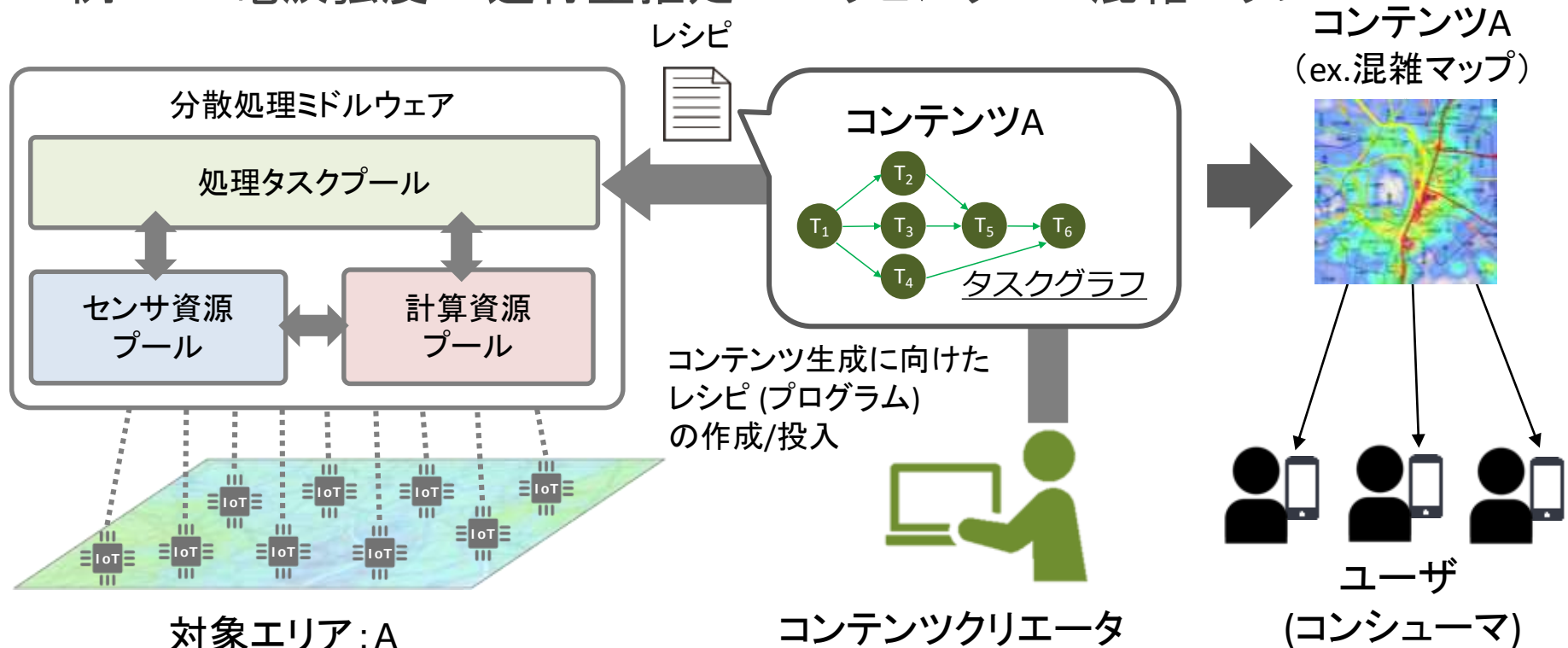
- タスクの要件を満たすように動的に処理クラスタを形成



技術的課題2: コンテンツ生成に向けたタスクをどう管理？

➤ クリエータによるコンテンツ毎のタスクグラフ記述/登録

- センサデータ → コンテンツに変換する処理をDAG(有向非巡回グラフ)として記述
- 例: BLE電波強度 + 通行量推定 + マッピング → 混雑マップ





何が必要か？

高速グラフィマイニング技術

グラフマイニング技術に期待すること



- 大規模に使える
 - 分析対象の巨大化
- 動的に使える
 - 時系列変化
- 簡単に使える
 - 操作性, PCパワーを必要としない