Winter Festa Episode 2

- 若手研究者主体のイベント

  → 大学院生・PD中心（25－35歳）

  ビジネス色はなるべく少なく、基礎技術、基礎研究中心

- 情報系（基礎）の大型プロジェクト結集！

  → 現在の日本の主要プレーヤーも集結！　研究はどれも世界レベル！

- タイムリーな招待講演！

  → JST＆文科省からの叱咤激励＆クリスマスプレゼント？

幅広い情報系基礎研究を披露＆若手研究者の交流

プロジェクトの紹介を少なめ（＜10分）＆ポスター中心（交流できるように！）

→ 若手研究者の方、ぜひ楽しんで下さい！

2016 年 12 月 22 日（木）
13:00-18:30（受付開始 12:30～）
2016 年 12 月 23 日（祝・金）
10:00-17:00（受付開始 9:30～）
一橋講堂
2F
中会議場
EPISODE 2
情報系 WINTER FESTA
主催
JST ERATO
河原林巨大グラフ
プロジェクト
共催
理化学研究所
革新知能統合
研究センター
（AIPセンター）
PROJECT
JST ERATO 河原林巨大グラフプロジェクト
革新知能統合研究センター（AIPセンター）
基盤（S）離散構造処理系プロジェクト
新学術領域研究 渡辺プロジェクト（ELC）
CREST 佐久間プロジェクト
CREST 宇野プロジェクト
CREST 加藤プロジェクト
CREST 津田プロジェクト
CREST 岩田プロジェクト
CREST 平岡プロジェクト
国立情報学研究所

Program

Day1　Thursday, December 22

13:00 ― 13:05	開会･主催者挨拶 河原林 健一 JST ERATO 河原林巨大グラフプロジェクト 研究総括
13:05 ― 13:30	ERATO 河原林プロジェクト紹介 / ハイライト 河原林 健一 （国立情報学研究所）
13:30 ― 13:45	CREST 岩田プロジェクト紹介 / ハイライト 岩田 覚 （東京大学）
13:45 ― 14:30	ポスターセッション
14:30 ― 14:55	CREST 佐久間プロジェクト紹介 / ハイライト 佐久間 淳 （筑波大学）
14:55 ― 15:10	CREST 平岡プロジェクト紹介 / ハイライト 平岡 裕章 （東北大学 原子分子材料科学高等研究機構）
15:10 ― 15:25	さきがけ研究者 ハイライト
15:25 ― 16:15	ポスターセッション
16:15 ― 16:45	新学術領域研究 渡辺プロジェクト(ELC)紹介 / ハイライト 渡辺 治 （東京工業大学）
16:45 ― 17:00	ACT-I研究者 ハイライト
17:00 ― 17:50	ポスターセッション
17:50 ― 18:10	特別講演 「情報系若手研究者への日本政府から期待と支援」 栗原 潔 文部科学省研究振興局参事官(情報担当)付専門官
18:30 ― 20:30	懇親会

Day2　Friday, December 23

10:00 ― 10:25	革新知能統合研究センター(AIPセンター)紹介 / ハイライト 杉山 将 （理化学研究所）
10:25 ― 10:40	CREST 津田プロジェクト紹介 / ハイライト 津田 宏治 （東京大学）
10:40 ― 11:20	ポスターセッション
11:20 ― 11:45	CREST 加藤プロジェクト紹介 / ハイライト 加藤 直樹 （関西学院大学）
11:45 ― 11:55	さきがけ研究者 ハイライト
11:55 ― 12:35	ポスターセッション
12:35 ― 14:00	昼食
14:00 ― 14:20	ACT-I「情報と未来」紹介 後藤 真孝 研究総括（産業技術総合研究所） 河原林 健一 領域アドバイザー（国立情報学研究所） 湊 真一 領域アドバイザー（北海道大学）
14:20 ― 14:50	基盤(S) 離散構造処理系プロジェクト紹介 / ハイライト 湊 真一 （北海道大学）
14:50 ― 15:05	CREST 宇野プロジェクト紹介 / ハイライト 宇野 毅明 （国立情報学研究所）
15:05 ― 15:50	ポスターセッション
15:50 ― 16:10	ERATO 河原林プロジェクト ハイライト
16:10 ― 17:00	ポスターセッション
17:00	閉会の挨拶

JST　ERATO巨大グラフプロジェクト

国立情報学研究所
河原林　健一

JST・ERATO採択＆国際研究センター発足！

JST戦略的創造研究推進事業（ERATO型研究）における
平成24年度 新規研究領域および研究総括の決定について（9月28日）

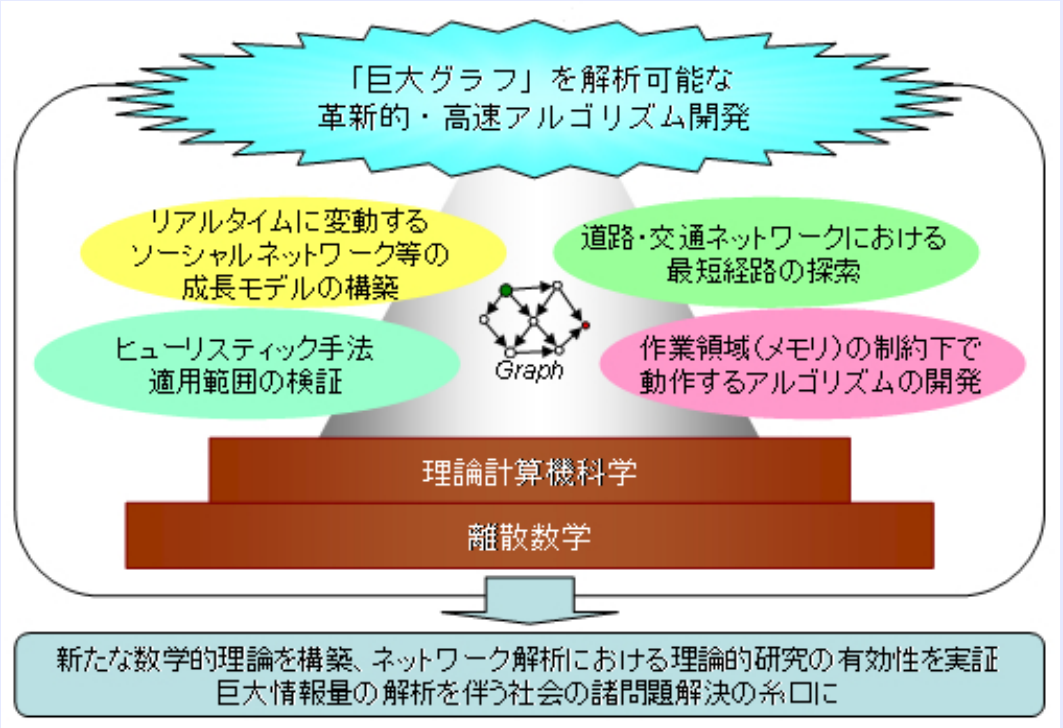
研究総括：河原林 健一（カワラバヤシ ケンイチ）（国立情報学研究所　教授）

研究領域：巨大グラフ

戦略目標：人間と調和する情報環境を実現する基盤技術の創出

研究期間：5年程度、研究費総額：最大12億円程度

研究内容：巨大なネットワークを解析する高速アルゴリズムの開発を通じて、革新的な数学的理論を構築し、大量情報の解析におけるその有効性の実証を目指します。



10月1日より、ERATO発足に伴い、NIIに「ビッグデータ数理国際研究センター」発足

もう4年目終了です！

- ～　2018年3月
- もうすぐ「審査」される側です！
- 目標、研究計画などはなく、あとは仕上げ！

ERATOでやったこと！

世界的な研究成果をあげるためには、

- 理論研究者を最低一人巻き込む
  - データホルダーも巻き込む
  - 現実的＆困難な課題を持っている人を巻き込む
- （実装力のある）大学院生＆PDが主導

↓

GOOGLE，MICROSOFT等のトップIT企業の研究所と同様の研究組織を目指す！（→このような研究所は日本では皆無！）

我々の研究成果('13-'16):数!

注意: 情報学分野では、論文誌よりトップ国際会議に論文が採用される方が重要!

- アルゴリズム&理論(STOC, FOCS, SODA) 34
  + SODA'13 &15, FOCS'15, Best Paper Awards!
- データベース (SIGMOD, VLDB, ICDE) 6
- 機械学習等(ICML, NIPS, AAAI, IJCAI) 29
- データマイニング(KDD, WWW, CIKM,WSDM) 13
- その他(SOCG, LICS, INFOCOM,CCC,ACL) 8

We had almost no experience before!

評価

1. 合計90本(以上)は, 日本全体の1/4~1/3.
2. 計算機科学分野の広範囲をカバーしている
3. ほぼすべての論文が,「数理科学」の理論研究ベース(多くの理論の組み合わせ)!
4. ほぼすべての論文で, 大学院生、ポスドクが主体!
5. 日本のVISIBILITY向上(1%程度)にも貢献
6. アルゴリズム分野でも質(3 BEST PAPERS)&量ともに向上!
7. 同じ予算規模の中では、世界の有数の研究センターに!

「純粋」な底上げ

	ERATO前		ERATO中
データベース (SIGMOD, VLDB, ICDE)	0	→	6
機械学習等(ICML, NIPS, AAAI, IJCAI)	4	→	29
データマイニング(KDD, WWW, CIKM,WSDM)	1	→	13
その他(LICS, INFOCOM,CCC, ACL)	0	→	5
合計	5	→	53

毎年　＋13～15　の底上げ！

コミュニティに訴えることができたこと
成功した研究例は…

1. 多くの大学院生・PDが主導

2. 「理論ベース」のアプローチ

3. 幅広い理論の道具を結集！

（若手研究者の）研究に必要な能力 ［重要度順］

1. アルゴリズムの理解
2. （組合せ）最適化の理解
3. 実装力
4. 機械学習・AI・データベース・データマイニングでの知識

人材育成

RA，PDの過去3年間でのポジション取得状況

（⇒ERATO終了までに1．5倍程度の規模になる可能性あり）

- トロント大助教(Assistant Professor)
- NII助教（2人）
- 東大助教（5人）
- 京大助教
- 東工大助教
- 大阪大助教
- 東北大助教
- 静岡大助教
- 豊橋技術大学助教
- NICTテニュアトラック研究員
- 産総研AIセンター
  テニュアトラック研究員（2人）
- 鳥取大助教
- 帝京大専任講師
- JSTさきがけ研究員

その他： NEC研究所、日立研究所、富士通研究所、Google、リクルートなど多数

RAから学振（DC）　→　14人

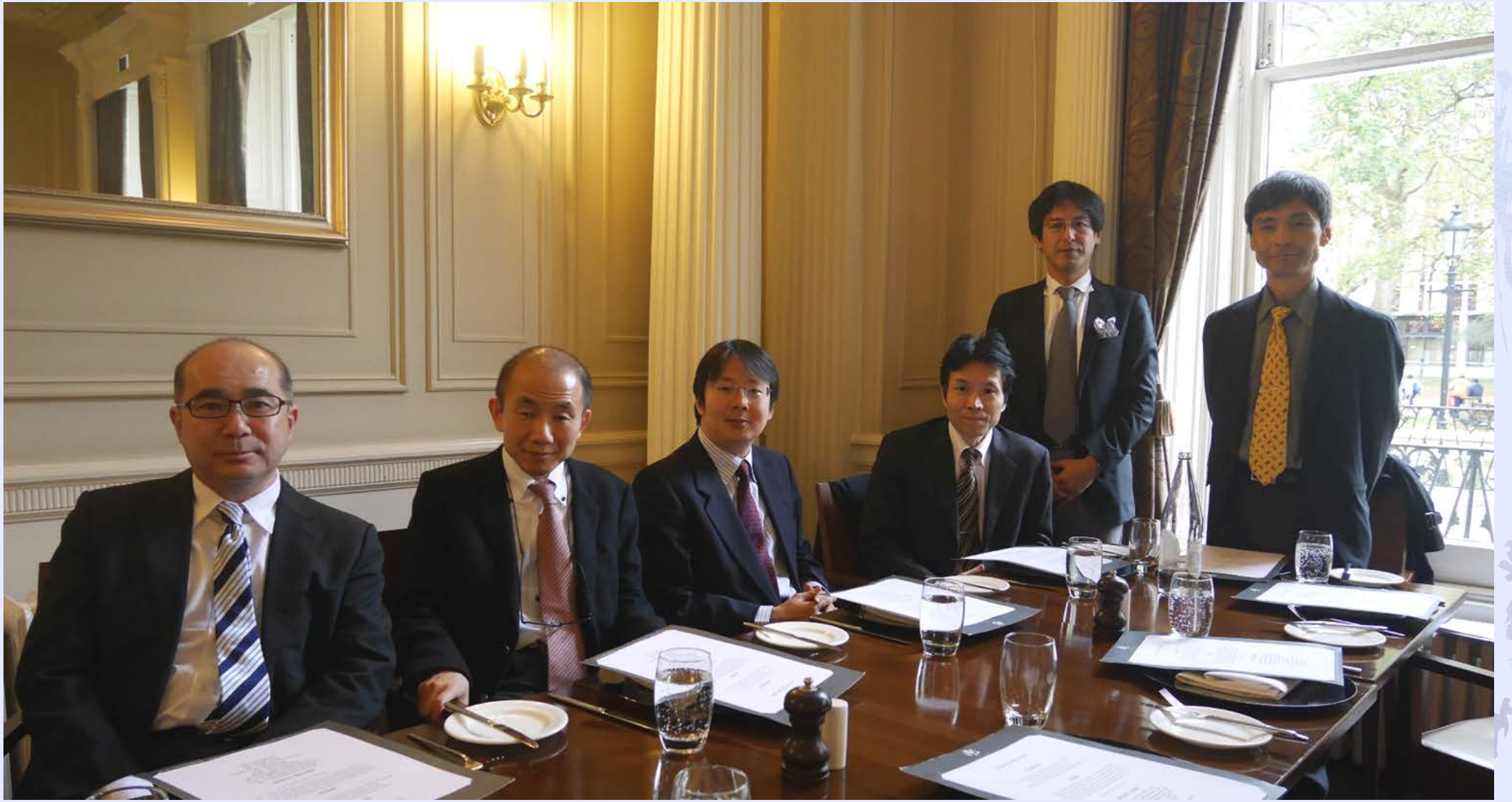
ERATOでコミュニティへのENHANCE

安倍首相のインド訪問に合わせて開催された「科学技術関連セミナー」

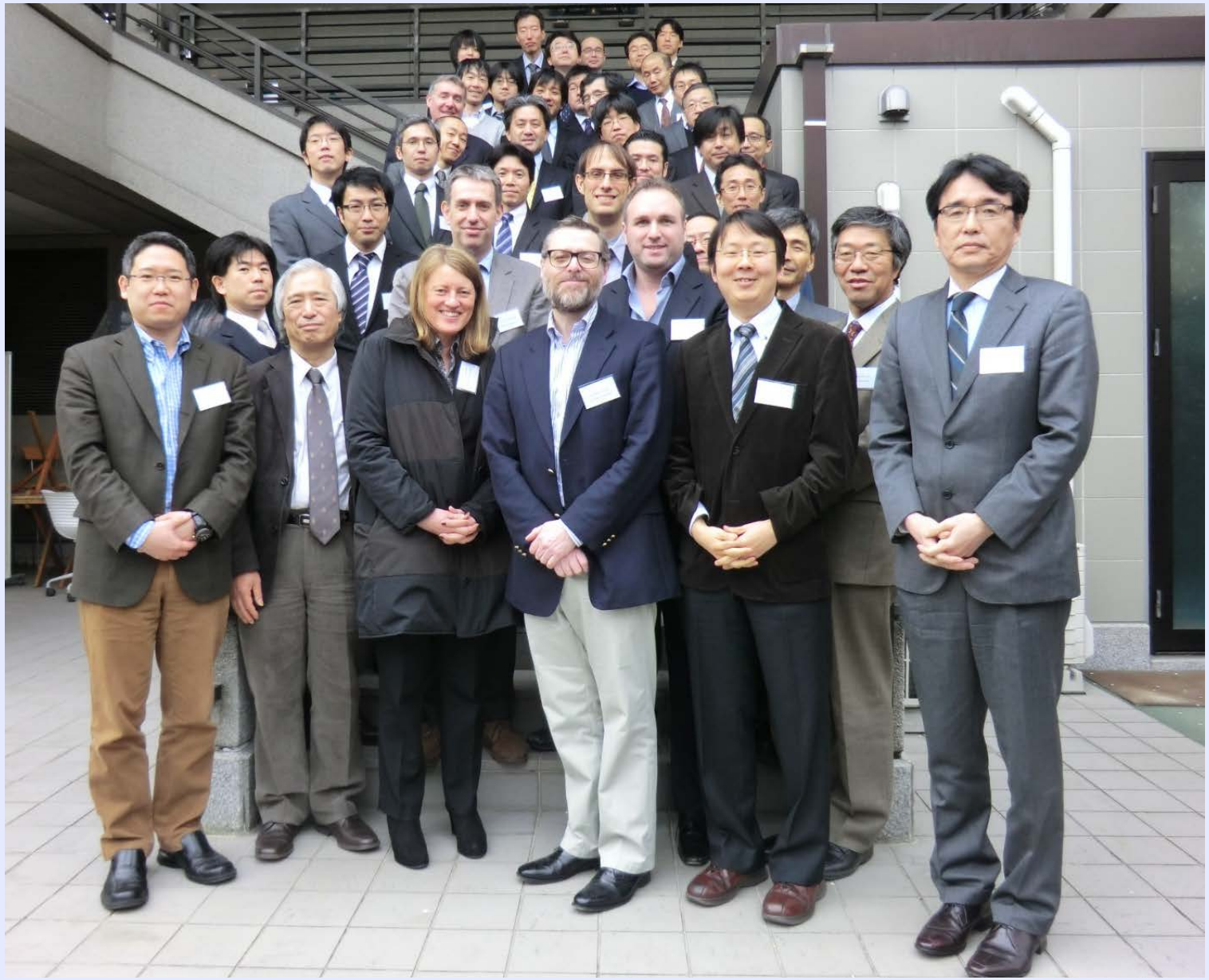
「科学技術関連セミナー」

FICCI
ogy and
ange"
ss and Academia (STSBA)
New Delhi
JETRO
JST
BIG DAT
Our background
BIG DATA

日英BIG DATA交流主催(英国訪問2014)



日英BIG DATA WORKSHOP主催(2014)



日仏BIG DATA WORKSHOP主催(2014)

日英BIG DATA交流主催(2015)
ウィリアム王子来日記念！

感謝祭

注意： 情報学分野では、論文誌よりトップ国際会議に論文が採用される方が重要！

- アルゴリズム＆理論（STOC, FOCS, SODA） 34
  + SODA‘13 &15, FOCS’15, Best Paper Awards!
- データベース (SIGMOD, VLDB, ICDE) 6
- 機械学習等(ICML, NIPS, AAAI, IJCAI) 29
- データマイニング(KDD, WWW, CIKM,WSDM) 13
- その他(SOCG, LICS, INFOCOM,CCC,ACL) 8

We had almost no experience before !

- このような研究成果は、ほとんど日本では発表していない！
  → 日本でも発表の場が必要！

- 本PJ外の研究者もトップ会議論文発表！
  → コミュニティに、現在日本で世界的な仕事を行っている（若手）研究者のショーケースを示す！

今年度の様子

Winter Festa

我々の組織は

離散数学，アルゴリズム，機械学習，データマイニング，自然言語処理，統計物理学，データベース，最適化，OR等，広範囲の研究者集団！

- **「数理科学」のキーワードのもとに数学，OR，情報科学のトップ研究者が集結**

⇒　情報処理・電気通信・応用数理・物理・応用物理・数学・ORなどの学会に広がる組織

⇒　学会の枠を超えた研究集会が必要！

クオリティの高い！情報系大型プロジェクト間の交流を行う！

INFORMATION
2016年12月22日(木)
13:00-18:30(受付開始 12:30~)
2016年12月23日(祝・金)
10:00-17:00(受付開始 9:30~)
一橋講堂
2F
中会議場
EPISODE 2
情報系WINTER FESTA
PROJECT
JST ERATO 河原林巨大グラフプロジェクト
革新知能統合研究センター
科研費湊課題
CREST 宇野課題
CREST 加藤課題
CREST 佐久間課題
CREST 津田課題
CREST 平岡課題
新学術領域研究渡辺研究領域
主催
JST ERATO
河原林巨大グラフ
プロジェクト
共催
理化学研究所
革新知能統合
研究センター
国立情報学研究所

Program （暫定版）

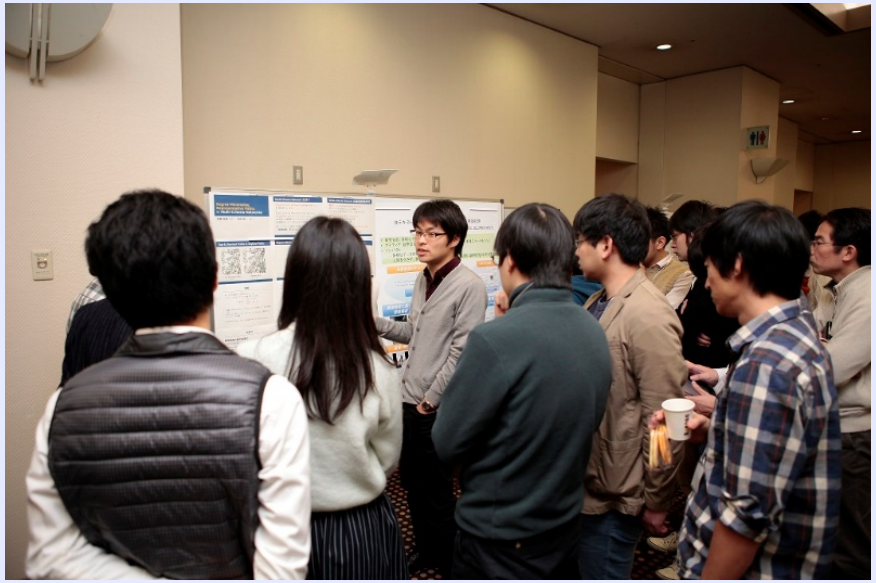
Day1　Thursday, December 22

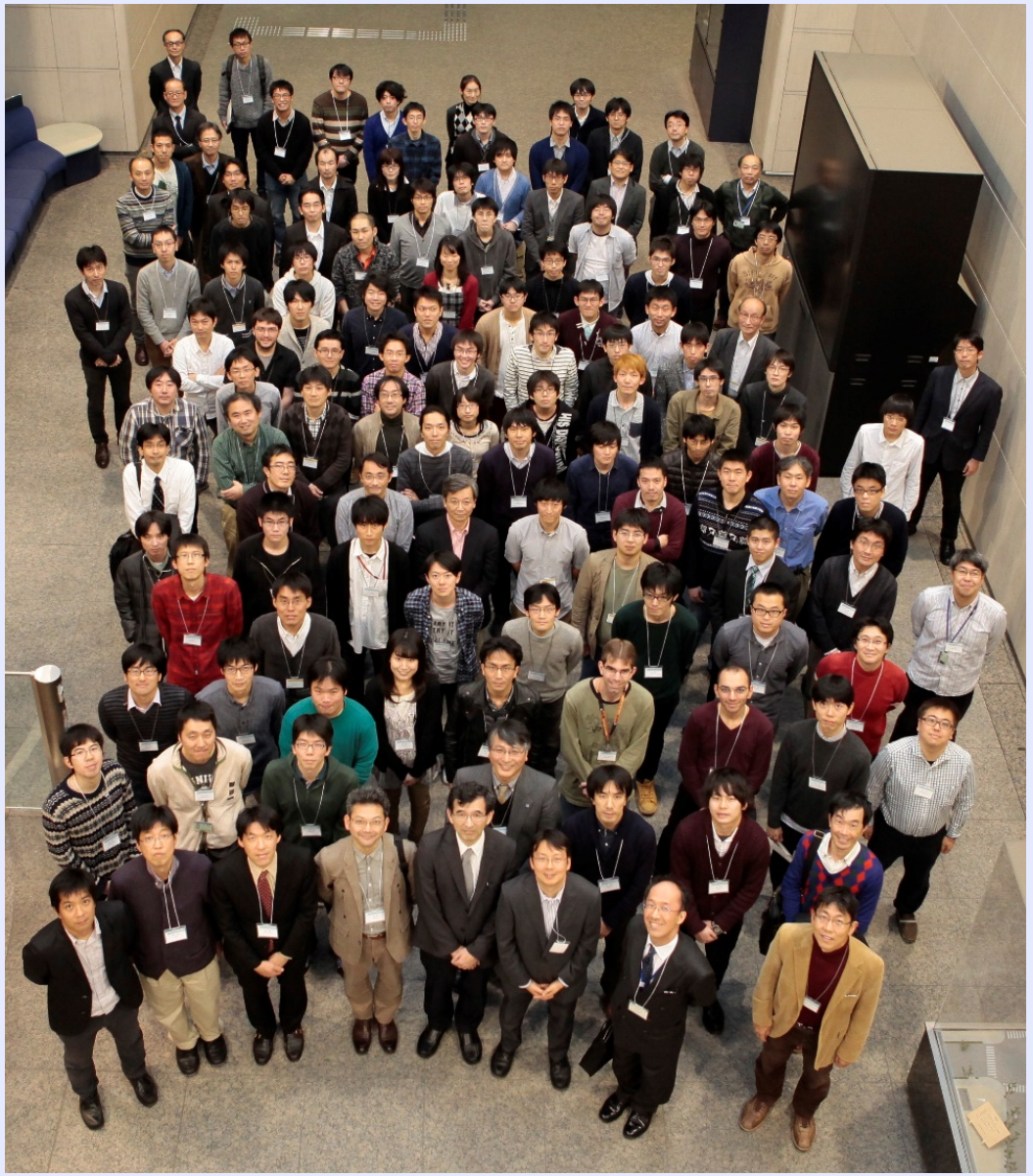
13:00 — 13:05	開会･主催者挨拶 河原林 健一 JST ERATO 河原林巨大グラフプロジェクト　研究総括
13:05 — 13:30	ERATO 河原林プロジェクト紹介　/　ハイライト_Group1 河原林 健一 （国立情報学研究所）
13:30 — 13:45	CREST 岩田プロジェクト紹介　/　ハイライト 岩田 覚 （東京大学）
13:45 — 14:30	ポスターセッション
14:30 — 14:55	CREST 佐久間プロジェクト紹介　/　ハイライト 佐久間 淳 （筑波大学）
14:55 — 15:10	CREST 平岡プロジェクト紹介　/　ハイライト 平岡 裕章 （東北大学　原子分子材料科学高等研究機構）
15:10 — 16:00	ポスターセッション
16:00 — 16:30	新学術領域研究 渡辺プロジェクト紹介　/　ハイライト 渡辺 治 （東京工業大学）
16:30 — 16:45	さきがけ・ACT-I研究者　ハイライト
16:45 — 17:30	ポスターセッション
17:30 — 18:15	ディスカッション
18:15 — 18:30	写真撮影
18:30 — 20:30	懇親会

Day2　Friday, December 23

10:00 — 10:25	革新知能統合研究センター紹介　/　ハイライト 杉山 将 （理化学研究所）
10:25 — 10:40	CREST 津田プロジェクト紹介　/　ハイライト 津田 宏治 （東京大学）
10:40 — 11:20	ポスターセッション
11:20 — 11:45	CREST 加藤プロジェクト紹介　/　ハイライト 加藤 直樹 （関西学院大学）
11:45 — 11:55	ERATO 河原林プロジェクト　ハイライト_Group2
11:55 — 12:35	ポスターセッション
12:35 — 14:00	昼食
14:00 — 14:30	基盤(S) 離散構造処理系プロジェクト紹介 / ハイライト 湊 真一 （北海道大学）
14:30 — 14:45	CREST 宇野プロジェクト紹介　/　ハイライト 宇野 毅明 （国立情報学研究所）
14:45 — 15:30	ポスターセッション
15:30 — 15:50	さきがけ研究者･ERATO 河原林プロジェクト　ハイライト_Group3
15:50 — 16:30	ポスターセッション
16:30 — 17:00	ディスカッション
17:00	閉会の挨拶

昨年度開催の様子





最後に

ERATO後（2018，2019以降）は？

白紙！

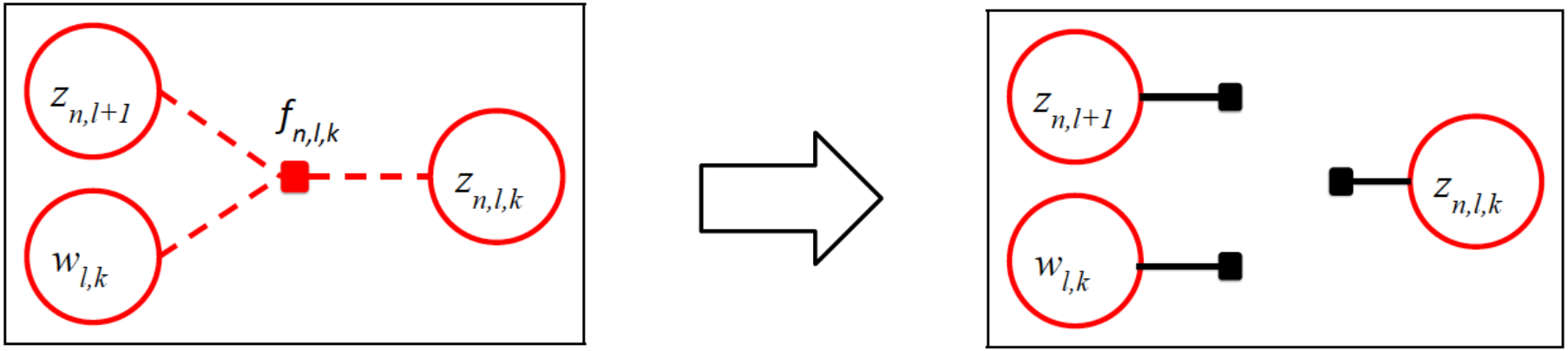
ポスター(ERATO Part 1)

Emtiyaz Khan (AIP)
波多野 大督 (NII&ERATO)
薗部 知大 (NII&ERATO)
小西 卓哉 (NII&ERATO)
原 聡 (NII&ERATO)
照山 順一 (NII&ERATO)
阿部 真人 (NII&ERATO)
関口 卓也 (NII&ERATO)
Man-Kwun Chiu (NII&ERATO)

Inference on Complex Bayesian Models using Simple Computations.

Mohammad Emtiyaz Khan and Wu Lin (AIP, RIKEN)

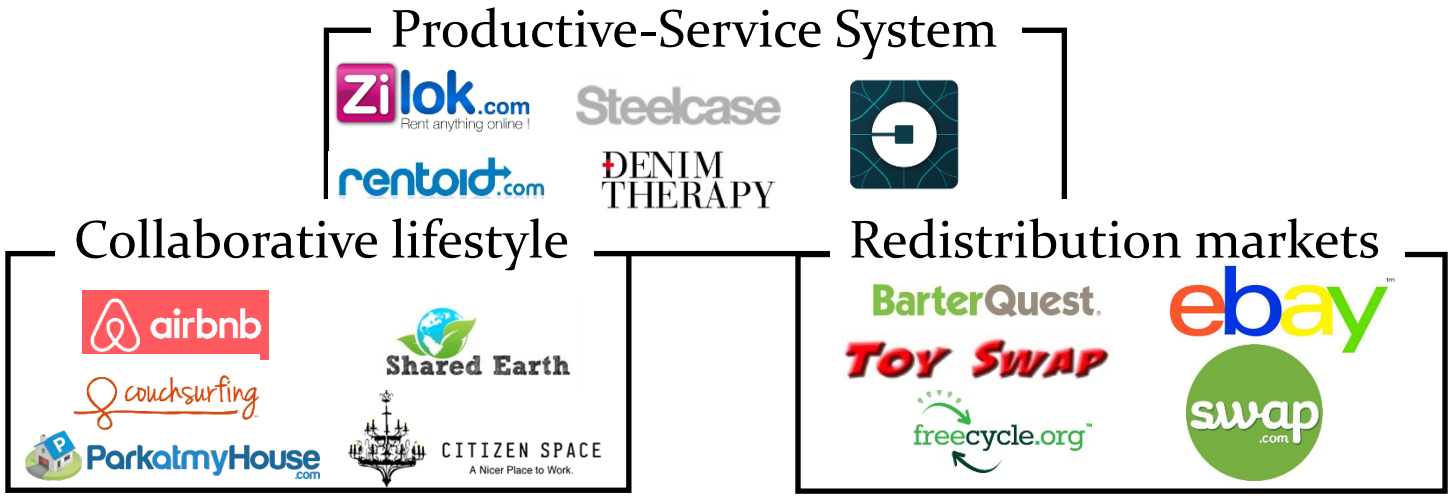
- New method for inference probabilistic graphical models that contain difficult terms by using Stochastic Message-Passing.
  - Non-conjugate terms (e.g. in logistic regression, we have Gaussian distribution with logistic likelihood).
- Main Idea: conjugate approximations for non-conjugate terms by using "proximal-gradient" method.
  - E.g. approximate logistic likelihood by Gaussian, which converts logistic regression to linear regression.
- Benefit: Modular, scalable, generalizable, and convergent updates.



Computing Least Cores of Supermodular Cooperative Game [AAAI 2017, to appear]

Daisuke Hatano(Joint work with Yuichi Yoshida)

Motivation: Sharing economy



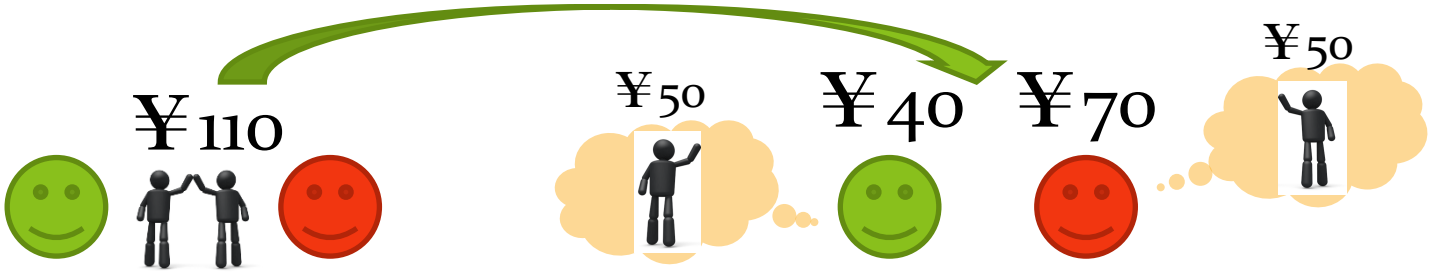
Need ***sharing*** for achieving the sustainable and resilient economy

Our goal: Find ***a value division*** to the players from which they have no incentive to deviate

Computing Least Cores of Supermodular Cooperative Game [AAAI 2017, to appear]

Daisuke Hatano(Joint work with Yuichi Yoshida)

Core: A solution concept of the game



Analyze the following solution concepts

- **Strong least core**
- **Weak least core**

Our contribution

Derive explicit and concise formulations for the strong and weak least core values

Looking Inside Literal Blocks: Towards Mining More Promising Learnt Clauses in SAT Solving

Tomohiro Sonobe

SAT instance

:

(a' v b' v x')

(b' v y')

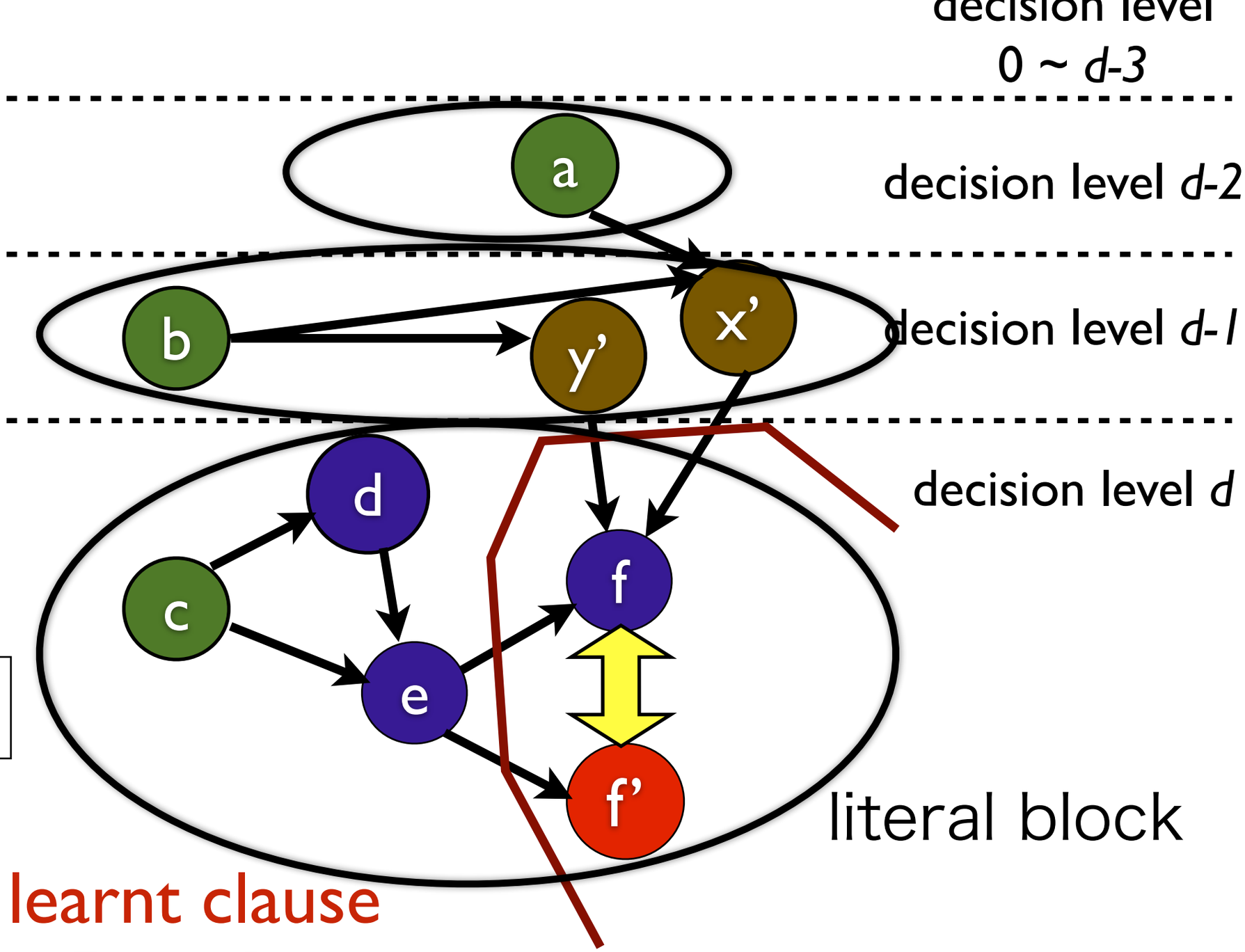
(c' v d)

(c' v d' v e)

:

(x v y v e' v f)

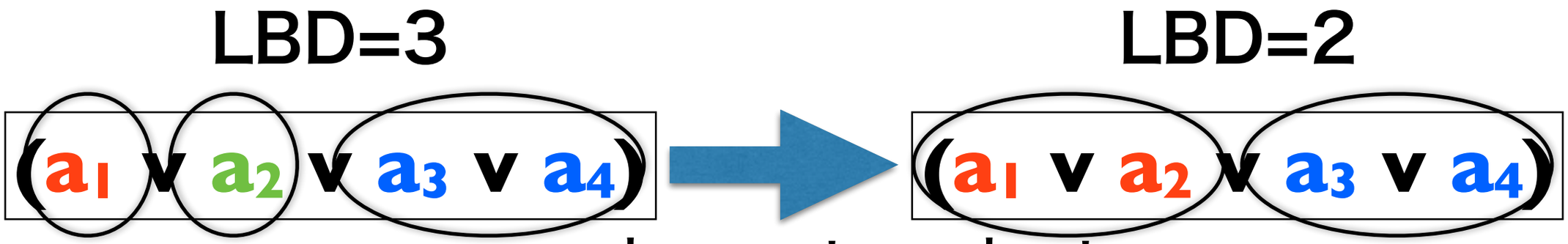
(e' v f')



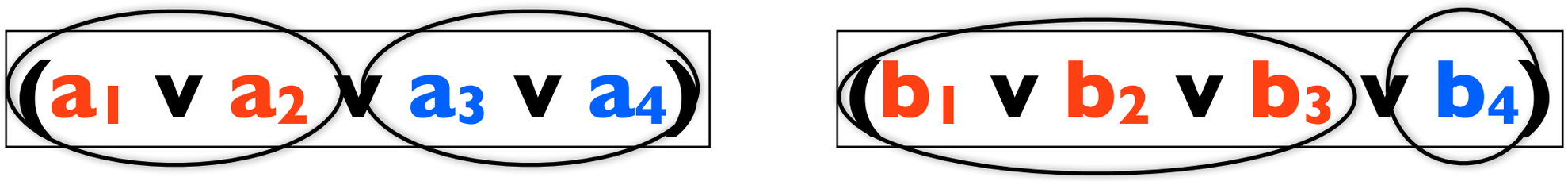
¬(x' ∧ y' ∧ e) <=> (x v y v e') Literal Block Distance (LBD)=2

Looking Inside Literal Blocks: Towards Mining More Promising Learnt Clauses in SAT Solving

Tomohiro Sonobe



- Learnt clauses with **potentially** low LBD value should be preserved.
- Learnt clauses with the same length and LBD are completely same?



Both clauses have length=4 and LBD=2.

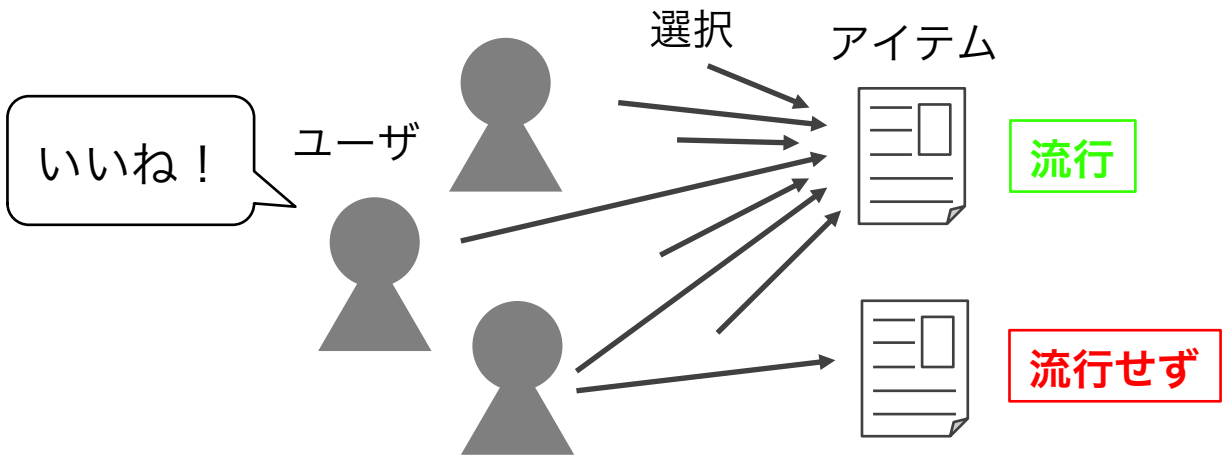
Features of literal blocks affect updates of LBD?

Identifying Key Observers to Find Popular Information in Advance

小西 卓哉（NII）　岩田 具治（NTT）
林 浩平（産総研）　河原林 健一（NII）

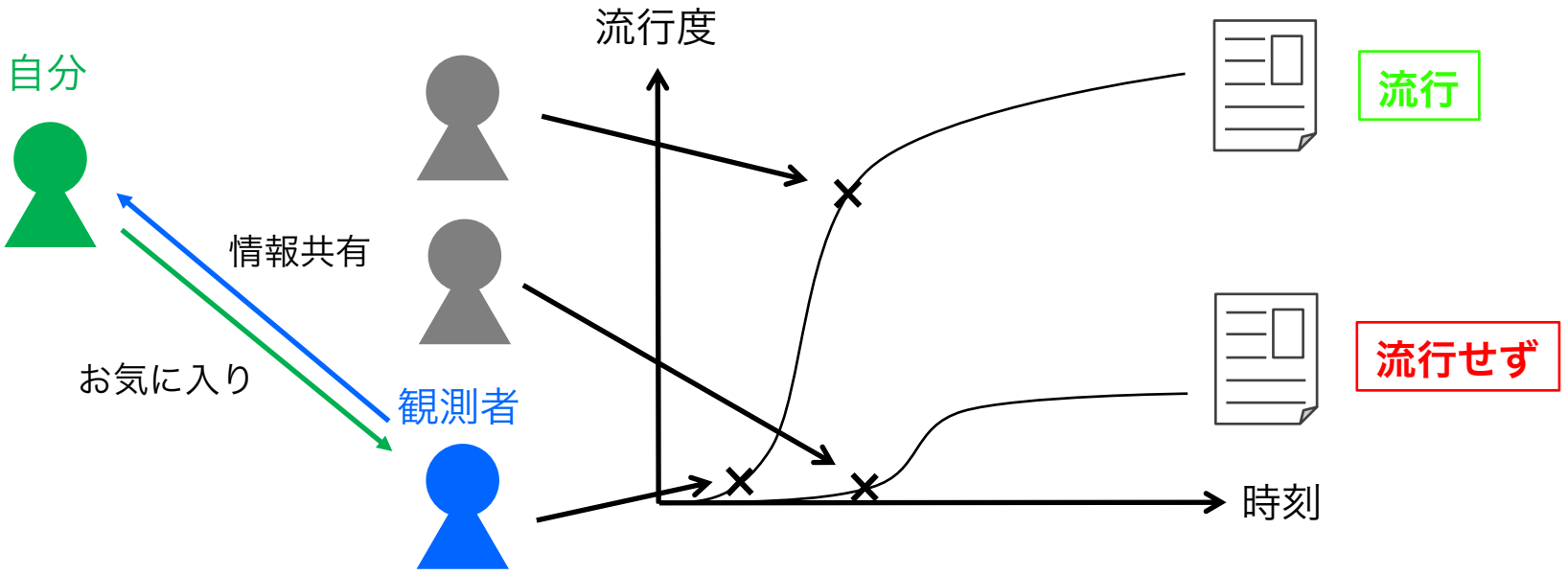
ウェブ上の情報共有と流行予測

- **アイテム（情報）**：メッセージ，写真，商品やサービス etc.
- **選択（アイテムの共有）**：リツイート，ブックマーク，レビュー etc.
- 将来流行しそうなアイテムを事前に察知したい



観測者を経由した流行の早期検知

- **観測者：エキスパート，ヘビーユーザなど**
- どのように有用な観測者をみつけるか？
  - **機械学習の特徴選択を応用して検出**

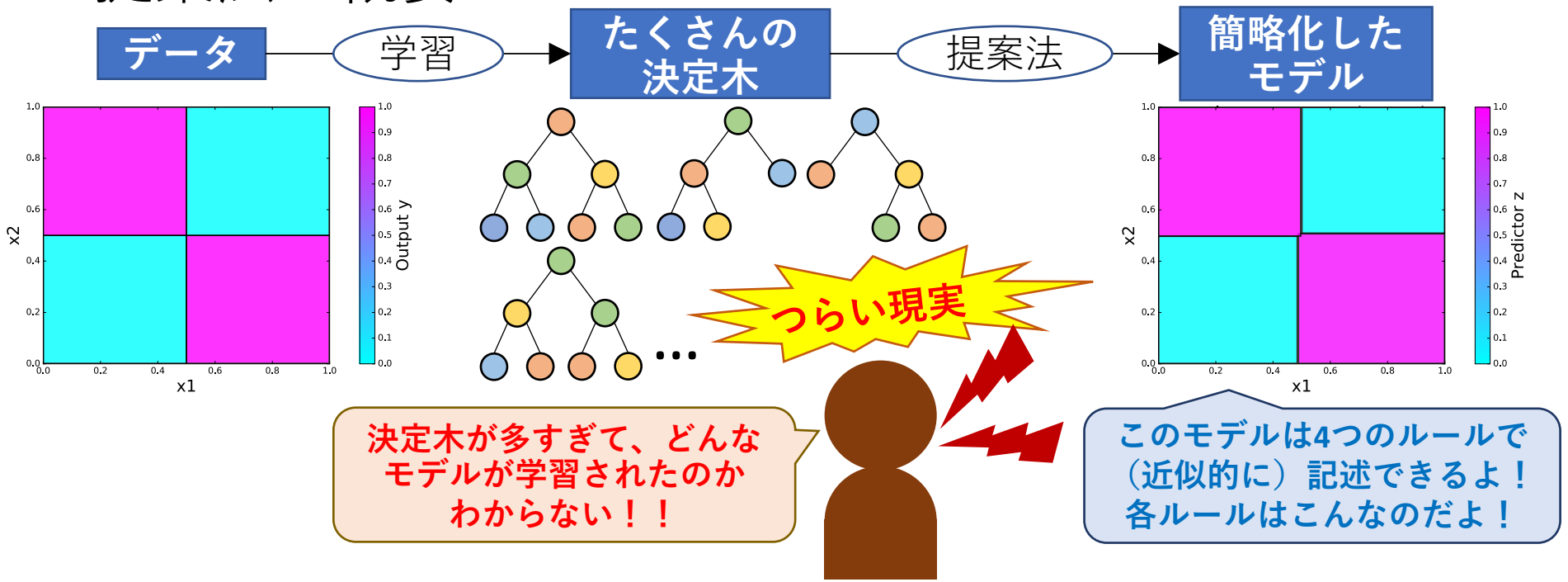


アンサンブル木モデル解釈のためのモデル簡略化法

原聡（NII/河原林巨大グラフプロジェクト）、林浩平（産総研）

- アンサブル木モデル：ランダムフォレスト、Gradient Boosted Treesなど
  - 長所：予測精度が高い。様々なデータに適用できる。
  - **短所：学習したモデルの解釈性が低い。**

- 提案法の概要



- 結果の一例：エネルギー効率データ
  - 各建物について8個の特徴量、768データ
  - 建物のエネルギー消費量の予測（回帰）

y	提案法 - 簡略化された予測ルール (K = 5)
12.14	Wall Area ＜ 306.25, Overall Height ＜ 5.25
14.41	Wall Area ≧ 306.25, Overall Height ＜ 5.25
27.80	Wall Area ＜ 330.75, Overall Height ≧ 5.25
28.45	Wall Area ＜ 330.75, Overall Height ≧ 5.25, Roof Area ≧ 116.375, Glazing Area Distribution ≧ 4
37.23	Wall Area ≧ 330.75, Overall Height ≧ 5.25

建物 - 小
→ 消費 - 小

建物 - 中
→ 消費 - 中

建物 – 大
→ 消費 – 大

- 小数のルールで小さいMSEを達成できた。

	ランダムフォレスト	提案法	inTrees（Deng, ‘14）	決定木
ルール数	＞1000	5	10	58
MSE	1.21	10.16	20.31	158.79

Satisfiability Algorithms for Read-k-times Branching Programs and k-IBDDs

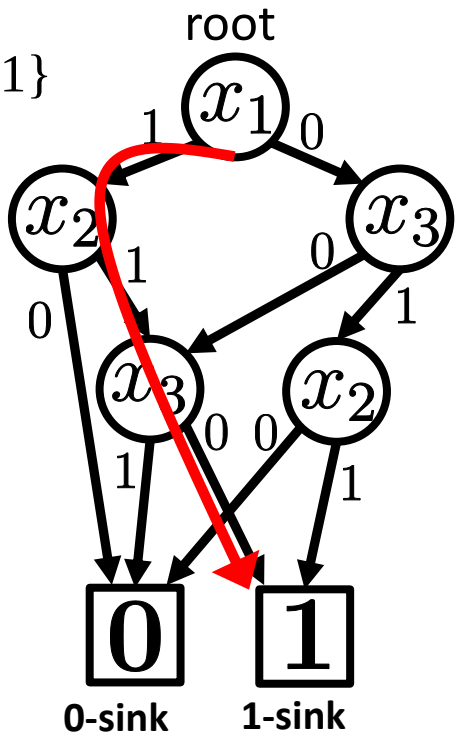
Junichi Teruyama (NII, JST ERATO)
joint work with Atsuki Nagao(Univ. of Electro-Communications), Kazuhisa Seto(Seikei Univ.)

Branching Program (BP)

- Computation Model
- Representation of $f : \{0, 1\}^n \to \{0, 1\}$

BP Satisfiability (BP SAT)

- Find an assignment
  that reaches to the 1-sink
- NP-Complete
- Exhaustive Search: $\mathbf{2^n}$ time
- Can we solve in time $\mathbf{1.99^n}$?
  exponential speed-up



x1	x2	x3	:	Output
0	0	0	:	1
0	0	1	:	0
0	1	0	:	1
0	1	1	:	1
1	0	0	:	0
1	0	1	:	0
1	**1**	**0**	**:**	**1**
1	1	1	:	0

super-polynomial speed-up

General

Time: $2^{n-n^{\delta}}, \delta = \delta(\epsilon)$, **Size:** $n^{2-\epsilon}$, **Space:** Exp.

[Chen, et al. (CCC2014)]

U

Read k-times

Result 1

Time: $2^{\left(1-\frac{1}{4^{k+1}}\right)n}$, **Size:** $\mathrm{poly}(n)$, **Space: Poly.**

exponential speed-up !!

UI

k-IBDD

Result 2

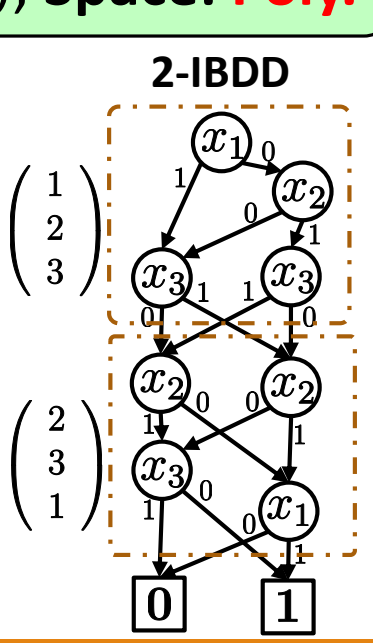
Time: $2^{\left(1-\frac{1}{2^{k}}\right)n}$, **Size:** $\mathrm{poly}(n)$, **Space: Poly.**

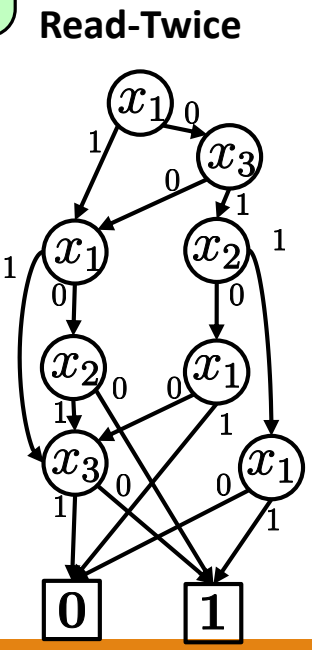
exponential speed-up for linear size input

Time: $2^{(1-\mu(c))n}, \mu(c) = \Omega\left(\frac{1}{\log^{2^{k-1}-1} c}\right)$

Input size: cn, **Space:** Exp.

[Nagao, Seto, **T.** (WADS2015)]





- Key Ideas
  - ✓ Decomposing Read-k-times BPs
  - ✓ Exhaustive Search for Disjoint Sets

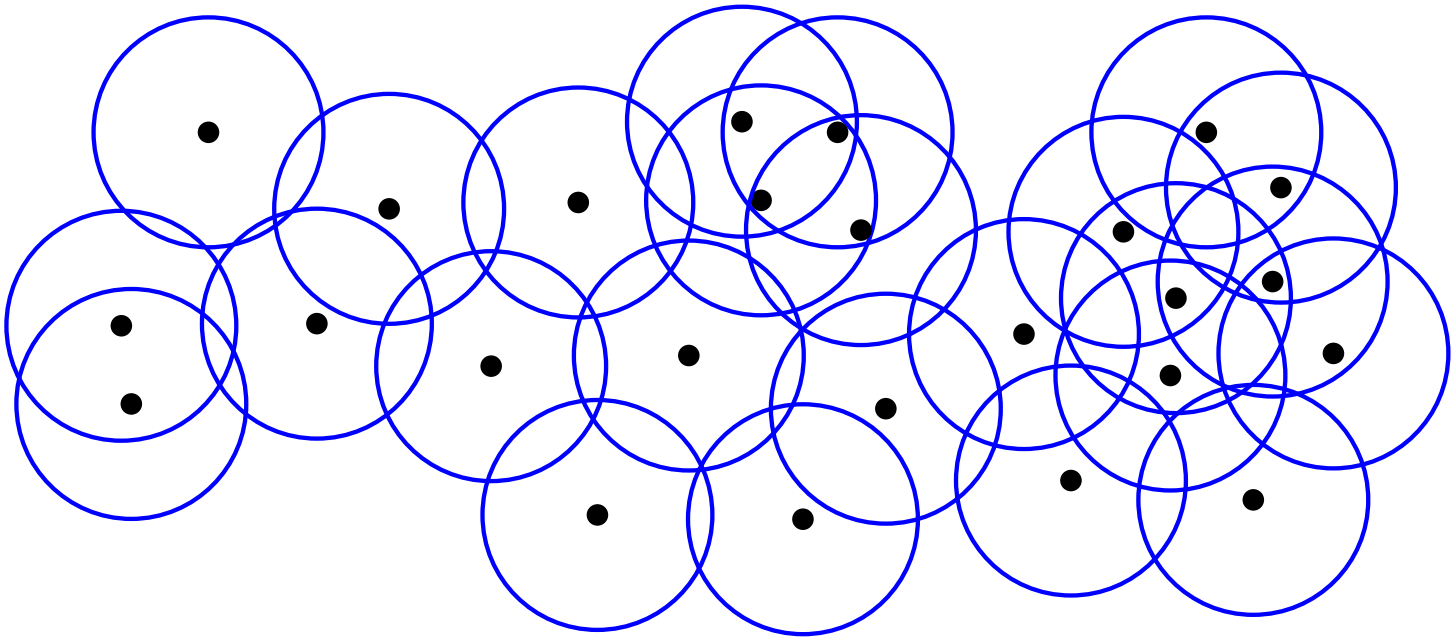
Line Separators of Unit Disk Graphs

Man-Kwun Chiu
chiumk@nii.ac.jp

National Institute of Informatics

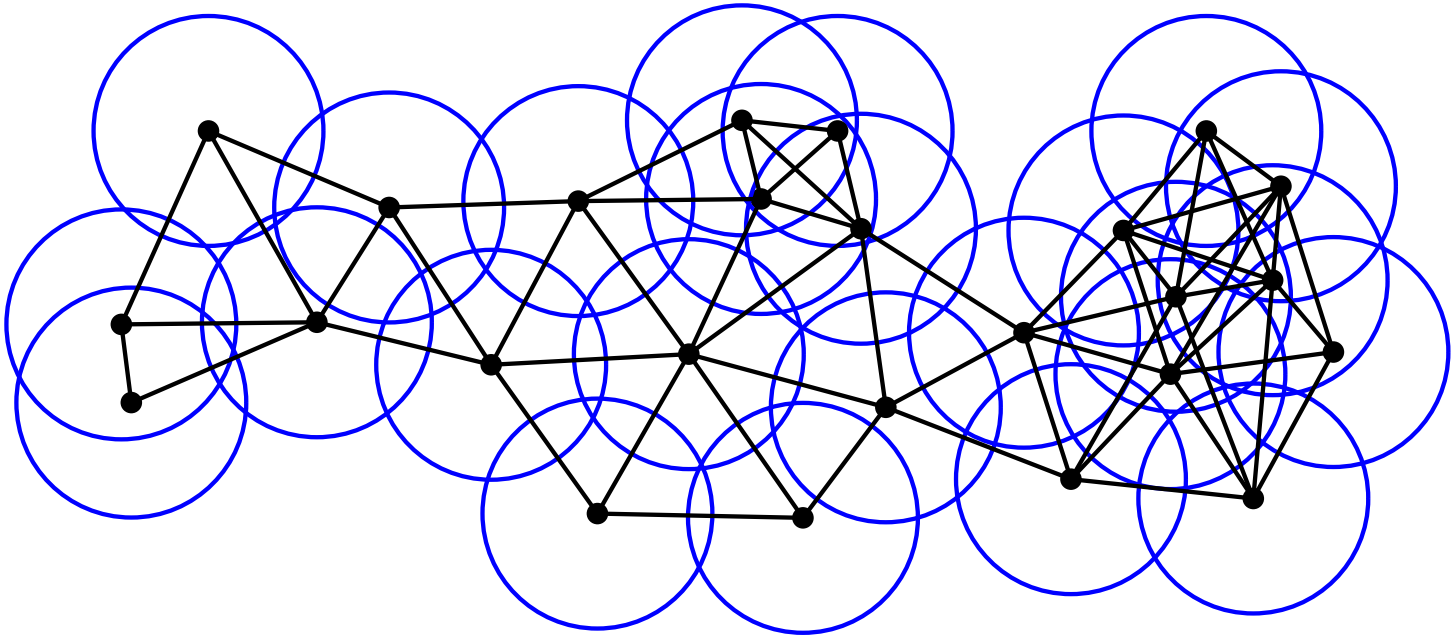
Joint work with Paz Carmi, Matthew J. Katz, Matias Korman, Yoshio Okamoto, André van Renssen, Marcel Roeloffzen, Taichi Shiitada and Shakhar Smorodinsky

Problem Definition - Line separator in Unit disk graph



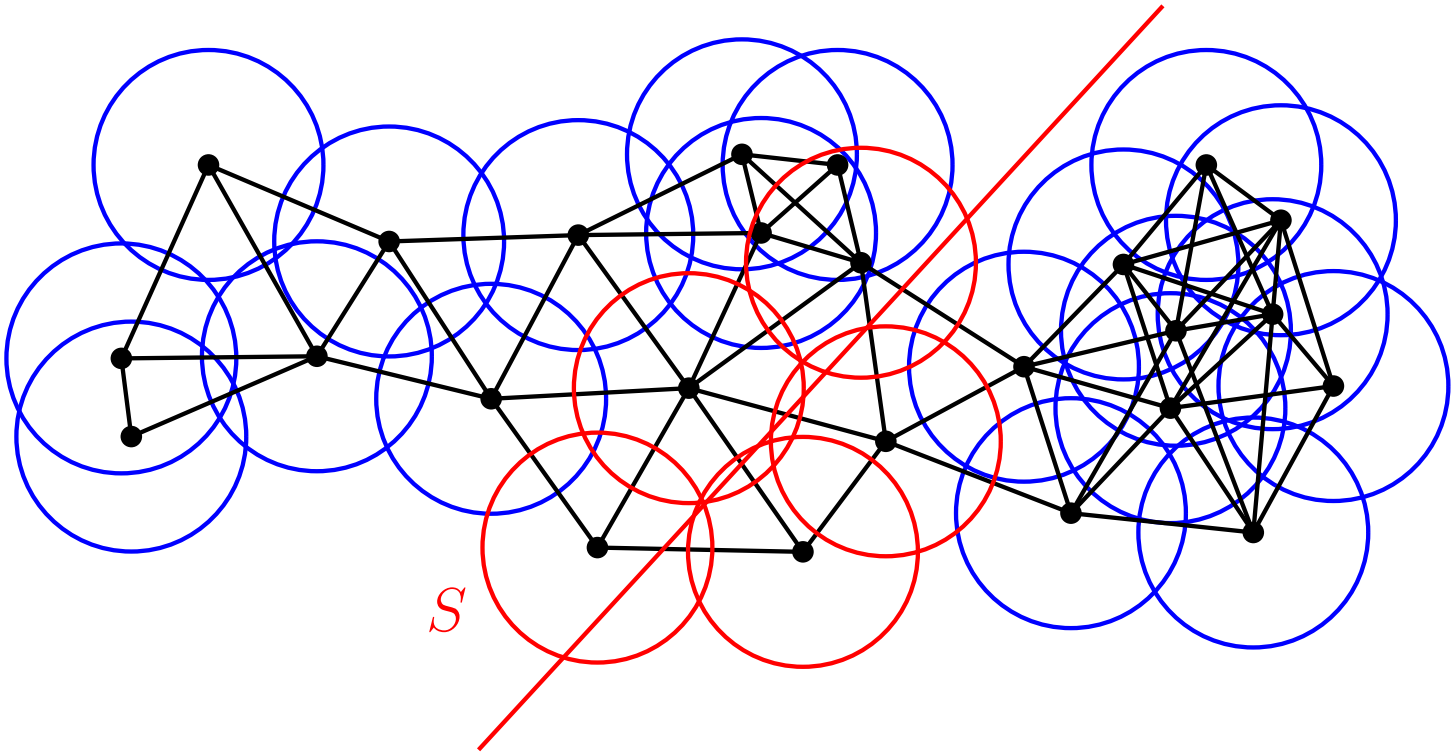
- Unit disk graph $G = (V, E)$:
  - Each vertex v_i represents a disk D_i.
  - $(v_i, v_j) \in E$ iff $D_i \cap D_j \neq \emptyset$.

Problem Definition - Line separator in Unit disk graph



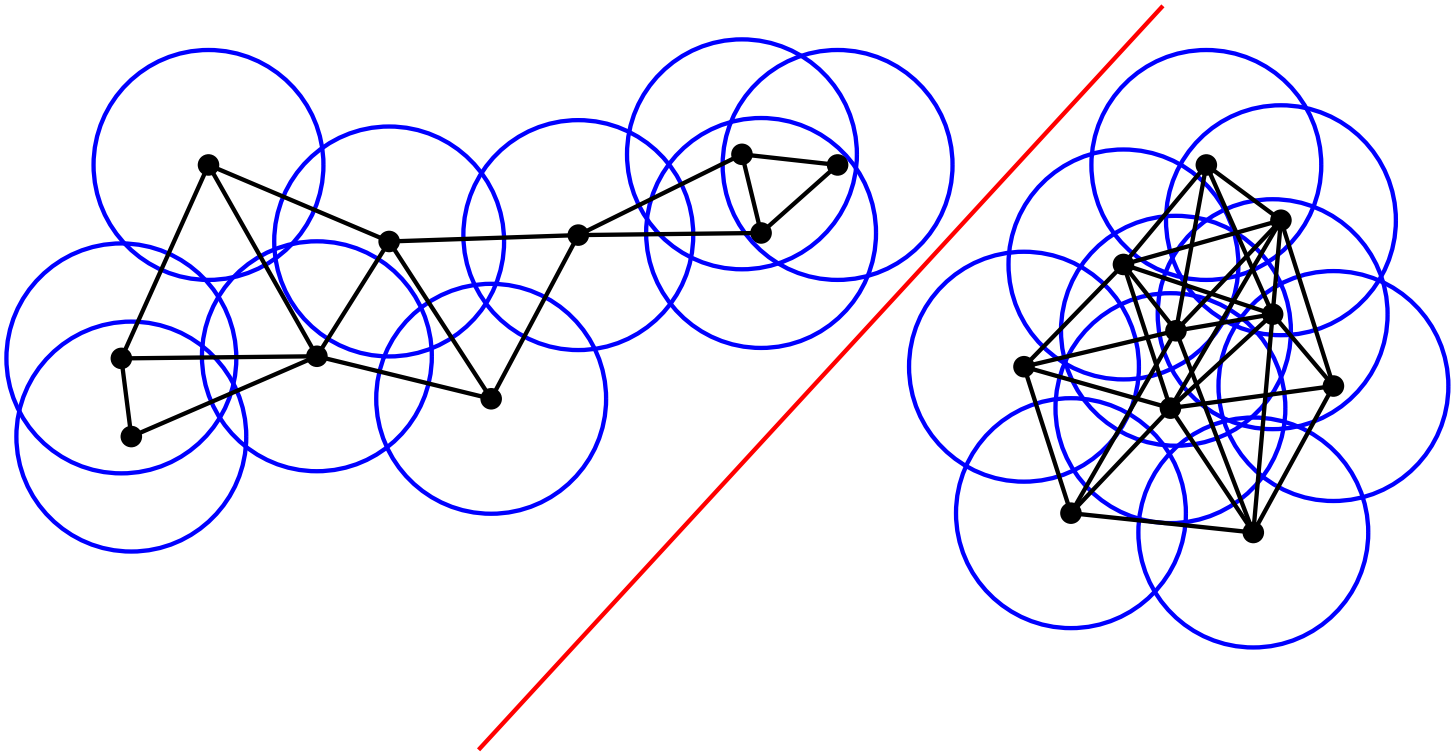
- A line α-separator $S \subseteq V$ is a set of all the disks intersected by a line ℓ such that each of two open halfplanes bounded by ℓ contains at most $\alpha|V|$ disks.
- Objective : minimize $\alpha \in [1/2, 1]$ and the size $|S|$.

Problem Definition - Line separator in Unit disk graph



- A line α-separator $S \subseteq V$ is a set of all the disks intersected by a line ℓ such that each of two open halfplanes bounded by ℓ contains at most $\alpha|V|$ disks.
- Objective : minimize $\alpha \in [1/2, 1]$ and the size $|S|$.

Problem Definition - Line separator in Unit disk graph



- A line α-separator $S \subseteq V$ is a set of all the disks intersected by a line ℓ such that each of two open halfplanes bounded by ℓ contains at most $\alpha|V|$ disks.
- Objective : minimize $\alpha \in [1/2, 1]$ and the size $|S|$.

Our Results

Upper bound on the size of the separator :

- For any unit disk graph with n disks and m intersecting pairs of disks, there exists a line 2/3-separator of size $O(\sqrt{(m+n)\log n})$.
- For any unit disk graph, there is an axis-parallel line 4/5-separator of size $O(\sqrt{m+n})$.

Lower bound on the size of the separator :

- For any unit disk graph, there does not exist a line* 2/3-separator of size better than $\Omega(\sqrt{m\log(n/\sqrt{m})})$.