

## 巨大グラフの解析とアルゴリズム

Analysis and Efficient algorithm for large graphs

河原林 健<sup>1\*</sup>

Ken-ichi Kawarabayashi

**概要** インターネットの Web 構造や, Facebook, Twitter などのソーシャルネットワークに代表される巨大なネットワークは, 各々  $10^9$  (10 億人) に近いユーザーが利用し, 現代社会に欠かせない存在となっている. このような背景のもと, JST ERATO 巨大グラフプロジェクトでは, 巨大なネットワークを膨大な点と辺の接続構造, すなわち「巨大グラフ」として表現し, 理論計算機科学や離散数学などにおける最先端の数学的理論を駆使してそれを解析する, 高速アルゴリズムの開発を目指している.

本講演では, 本プロジェクトの研究成果の一部を紹介する. 特に巨大グラフに対して, 理論研究をもとにできた効率的アルゴリズムをいくつか紹介する.

**キーワード** 巨大グラフ, グラフ解析, 高速アルゴリズム

---

1 国立情報学研究所, 〒101-8430 東京都千代田区一ツ橋 2-1-2

National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

\* E-mail address: k\_keniti@nii.ac.jp

# 巨大グラフの解析とアルゴリズム

## Analysis and Efficient algorithm for large graphs

国立情報学研究所  
JST ERATO 巨大グラフプロジェクト  
河原林 健一

# 巨大グラフに関する理論研究と実用研究の現状

情報爆発時代(ビッグデータ処理)を迎え、理論研究者へのニーズが高まっている  
理論研究の経験を持った意欲的な研究者が実用研究へ参入することで、大きな成果が収められつつある

- 理論の導入による新しい実用手法
  - ◆ PageRank・検索エンジン(Google)の最大の財産)
- 巨大ネットワークの解析は、グラフ理論的・離散アルゴリズム的解析が不可欠！←2011年DARPA(GPS,WEBを説明！)の研究テーマ
  - ◆ Tarjan, Karp, Hopcroft(チューニング賞受賞者)らも指摘！

理論分野で常に先端的な研究を行ってきた強みを生かし、上記の研究の流れの中で中心的役割を果たすことを目標とする

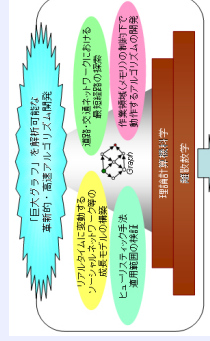
# Defense Advanced Research Projects Agency (DARPA)

- Funding Opportunity Title –**  
Graph-theoretic Research in Algorithms and the Phenomenology of Social networks (GRAPHS)
- Focused Areas**
- ◆ *The Structure and Dynamics of Large Networks*
    - Analysis of “Social Network”
    - Modeling “Social Network”
  - ◆ *Approximate Algorithm Development for Large Network*
    - ✓ GPSとインターネットを説明！
    - ✓ 基礎研究はサポートしない
    - ✓ 国家戦略にあった課題のみサポート！
    - ✓ 金額はINSEFの10倍以上！

# JST・ERATO採択&国際研究センター発足！

JST戦略的創発研究推進事業(ERATO創研究)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(カワハラケンイチ)(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礎技術の創出  
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度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。



新創発研究領域の構築、ネットワーク解析における理論的・実証的研究の推進、巨大情報の解析を伴った社会ネットワークの構築

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

## プロジェクトの目的

**究極的目標:** シヤノンのような「未来」のテクノロジーに革命を起こす「概念」を基礎研究(数学)に基づき提唱!

◆世界的に評価される基礎研究者で、かつ実社会への問題にも貢献できる人材を輩出する→未来への投資!

**現状:** GOOGLE, MICROSOFT, Amazon, Facebookなどの巨大IT企業の研究所の超一流の基礎研究者に水をあけられている!

## なぜ基礎研究が必要?

**現状:** GOOGLE, MICROSOFT, Amazon, Facebookなどの巨大IT企業の研究所の超一流の基礎研究者に水をあけられている!

なぜ?

1. 巨大IT企業が基礎研究の重要性を認識している(コンピューターのハード面の進歩のみではなく、理論研究に基づく改善が実際のサービス改善につながっている. Ex. PageRank)
2. 実データにアクセスできる → 「ビッグデータ」時代に必要不可欠

**日本の現状**

- 1の認識は非常に低い → **我々のセンターの目的!**
- 2はブライバシーの問題等 → 国家的課題!

## 研究例: Personalized PageRank

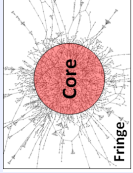
- ◆  $u \in V$  is important  $\Leftrightarrow u$  is frequently visited
- ◆ Visitors = Random Walkers
  - ◆ with prob  $\alpha$ , randomly moves to an adjacent vertex
  - ◆ Otherwise, randomly jumps to some vertex
- ◆ Personalized PageRank (PPR) = stable distribution:  
 $x = \alpha P x + (1 - \alpha) b$ 
  - $P$ : transition matrix
  - $b$ : jump target distribution
- ◆ Our task: compute PPR vector  $x \in R^n$

## Existing Methods for PPR

- ◆ Monte-Carlo Simulation (fast, inaccurate  $10^{-3}$ )
- ◆ Linear algebraic method (slow, accurate:  $10^{-9}$ )
  - ◆ **Direct methods:**
    - Gaussian elimination (LU decomposition)  
usually very slow for real world networks
    - ◆ **Iterative methods:**
      - Power iteration**  
state of the arts for PPR problem
      - Krylov subspace method**  
standard in numerical linear algebra,  
but not so effective for PPR problem

## Our Technique

- ◆ Decompose a network into **fringe** and **core**
  - ◆ fringe: tree-like subgraph
  - ◆ core: expander-like subgraph
- ◆ Apply **direct method** to fringe
  - ◆ It runs in linear time
- ◆ Apply **iterative method (GMRES)** to core
  - ◆ It converges very fast



(More precisely: we use the **inverse** of the fringe as a **preconditioner** of iterative method)

## Computational Results

- ◆ 1/5 iterations and 1/3 real-time than power iteration
- ◆ < 600 [s] for  $|V|=105M$ ,  $|E|=3.7B$  network
- ◆ In particular, effective on “web-graphs” (core is small → almost solved by direct

name	precond. GMRES		Power		naive GMRES	
	time[s]	iteration	time[s]	iteration	time[s]	iteration
in-2004	2.55	23	4.65	108	8.44	64
it-2004	136.08	21	321.65	104	572.33	74
uk-2007-05	576.98	26	1078.10	103	1626.06	73
twitter-2010	540.49	22	1824.23	106	1434.89	50
web-BerkStan	0.91	15	2.48	107	6.09	74
web-Google	1.88	27	2.40	107	4.91	64
web-NotreDame	0.18	17	0.61	102	0.98	71
web-Stanford	0.48	20	0.82	106	1.70	60
livejournal	21.20	21	69.11	106	35.79	37
flickr	3.50	18	11.35	108	7.18	37
orkut	41.69	21	120.33	74	46.99	24

## Conclusion

- ◆ proposed a fast and accurate PPR algorithm by exploiting **network structure** (fringe / core)
  - ◆ Apply **direct method (LU dec)** to the fringe
  - ◆ Apply **iterative method (Krylov)** to the core
- ◆ Beautiful combination of **Graph theory & Numerical Linear Algebra**
- ◆ Outperforms the existing methods in experiments

## 研究目標: アルゴリズムが世界を変える!

- ◆ PageRank, SimRankなどの世界的なアルゴリズム研究課題に対して, 高速化・スケール化を目標
- ◆ アルゴリズムの高速化によって新しいサービス提供へ(とくにWEBやSOCIAL NETWORK→Yahoo!との共同研究)
- ◆ プロジェクトの紹介ムービーを参照 (アルゴリズムが世界を変えるSeason1.0)

<https://www.youtube.com/watch?v=gZPMtEujEJc>  
(4日間で>1400 views!)

## 人材育成:研究体制

- 研究総括: 河原林健一 (NII)  
研究総括補佐: 今井和雄 (NII)
- 《研究規模》  
特任教員: 6名 PD: 15名前後 RA: 25名前後  
《年齢構成》  
> 50名: 1名 34~40歳: 1名  
28~33歳(助教, PD): 20名 < 28歳(大学院生): 25名
- 大学院生、PDが世界トップクラスの研究成果を残す!
  - MIT, Stanfordのエリート学生とも色んな色合い、若手研究者を10~15人供給!
  - 5~10年後に世界的拠点へ!

## センター活動予定

- ◆ 複雑ネットワーク・地図ネットワークグループ
- WEB & Social Networkの計算課題 (PageRank, 最短パスetc)
- Social Networkの社会心理学的分析
- WEB全体グラフ (500B辺以上)の解析 (Yahoo!との共同研究)
- Computational Advertising (Kakaku.com (電通) と共同研究)  
→ **Web上の広告価値の決定**
- Clique-Query (Yahoo!との共同研究)
- ◆ 理論・ネットワークグループ
- Cash配置問題 (SINETデータ利用の予定)
- 物理シミュレーションとアニーリング (日立と共同研究)

## 次のステージへ!

- ◆ 3~5年後には、世界に通用する若手研究者を**10~15**人輩出!
  - ◆ 若手研究者を中心に世界的な「科学的」そして「オリジナリティ」のある研究成果! 日本の「底上げ」&競争方向向上に貢献  
(情報系の大型PJのひとつのあり方を示す!)
  - ◆ この「オリジナリティ」のある成果をどうやって社会につなげるか?  
→ 今後の課題
- 必要なこと: エンジニアとの共同作業!**
- 世界最先端の研究成果と実社会の課題をつなぐ人材の供給
  - 「基礎研究」と「応用研究」をつなぐ人材の供給!