でたらめな配線がネットワークをサクサクにする

未来のスパコン／データセンターのネットワーク

Interconnection Technologies for Next-generation Supercomputers and Datacenters

藤原一毅，胡 曜，Truong Thao NGUYEN，鯉渕道紘　I. Fujiwara, Y. Hu, T.T. Nguyen, M. Koibuchi

どんな研究？

スーパーコンピュータの正体は、ネットワークでつながった数千台の小さなマシンです。マシンがどんなに速くても、ネットワークが遅かったらスパコンは性能を発揮できません。また、数千台のマシンをつなぐ、1,000キロメートルを超えるケーブルも悩みの種です。私たちは、スパコンのネットワークを「**もっと速く**」「**もっとスリムに**」する方法を探究しています。

何がわかる？

ネットワークの「低遅延化」と「省資源化」を両立するために、私たちは2つの方法を考えています。ひとつは、今まで規則的につないでいたマシン同士を**ランダムにつなぐ**こと。もうひとつは、ケーブルの代わりに**レーザービーム**を使うこと。これらの方法により、次世代のスパコンに求められる高性能ネットワークをシンプルに実現できることがわかりました。

状況設定

- 従来のスパコンのネットワークの限界
  - 計算速度と通信速度の乖離
  - 大きなスイッチ遅延と規則的トポロジ
    → 通信遅延が小さくならない
  - ケーブルだけで 100 トンを超える物量
    → 故障しても修理できない

広帯域指向／重厚長大な設計

© Fujitsu

▲ 3D Torusトポロジ（京コンピュータ）

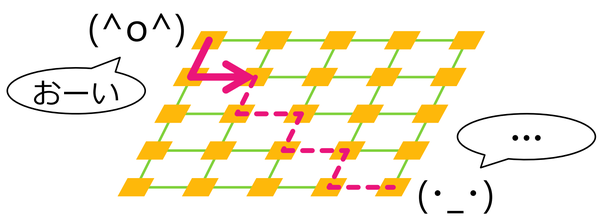
- 次世代スパコンに求められる
  ネットワークの要件
  - ノード数　～10,000
  - 通信遅延　1 マイクロ秒以下
  - メッセージサイズ　3KB 未満

小さなメッセージを／頻繁に／
遠いところへ／早く届けたい！

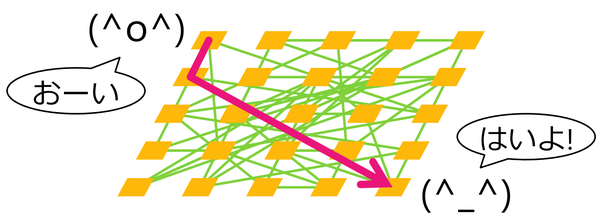
▲ 2,400kmに及ぶ配線（地球シミュレータ）

研究内容（1）

■ランダムトポロジ

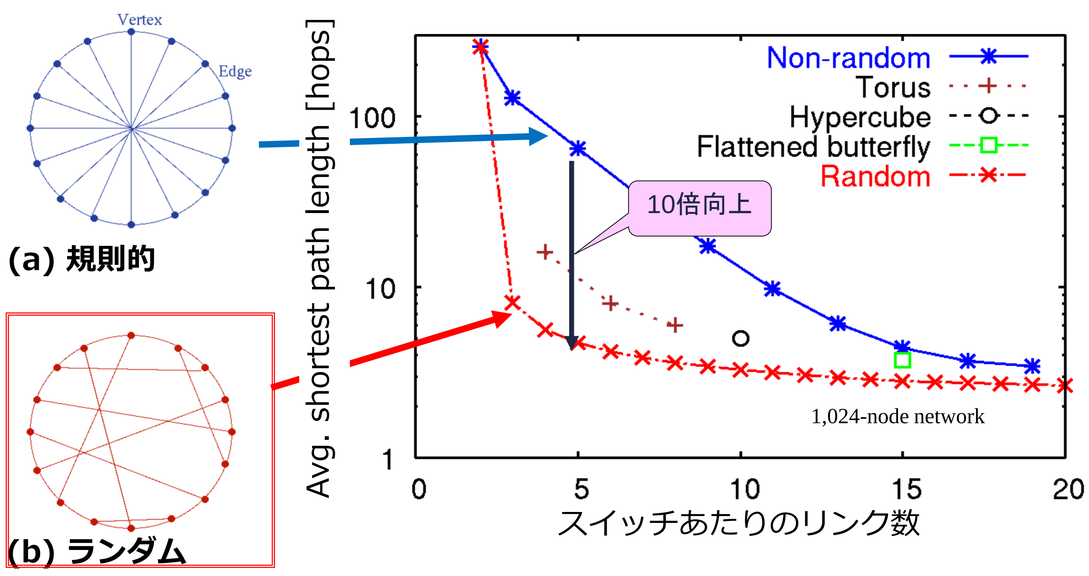


▲ 規則的なトポロジ：
メッセージが届くのに時間がかかる



▲ ランダムトポロジ：
近道があるのでメッセージがすぐ届く

- ノード間をランダムにつなぐと、スモールワールド効果により大きなネットワークほど劇的に距離が縮まり、遅延が減る [1]
- ただし、ケーブルが長くなる → ラック配置最適化で緩和 [3]
- 既存ネットワークのケーブルをランダムに差し替えるだけでも同様の効果 → ケーブルを引き直さずにランダムトポロジ化 [2]



発表論文

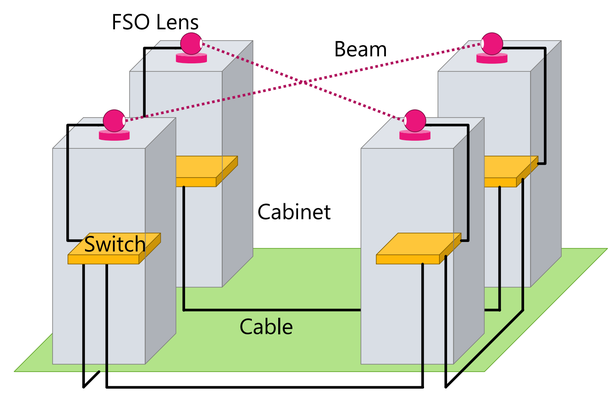
[1] Michihiro Koibuchi, Hiroki Matsutani, Hideharu Amano, D. Frank Hsu, Henri Casanova: "A Case for Random Shortcut Topologies for HPC Interconnects", The 39th International Symposium on Computer Architecture (ISCA), Jun. 2012.

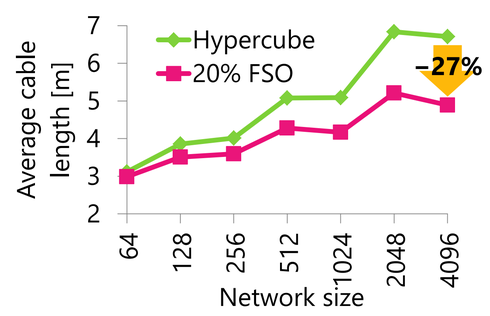
[2] Michihiro Koibuchi, Ikki Fujiwara, Hiroki Matsutani, Henri Casanova: "Layout-conscious Random Topologies for HPC Off-chip Interconnects", The 19th International Symposium on High-Performance Computer Architecture (HPCA), Feb. 2013.

[3] 藤原一毅, 鯉渕道紘: "ランダムなネットワークトポロジのラック配置最適化に関する研究", 電子情報通信学会論文誌 J96-D(8), 2013年8月.

研究内容（2）

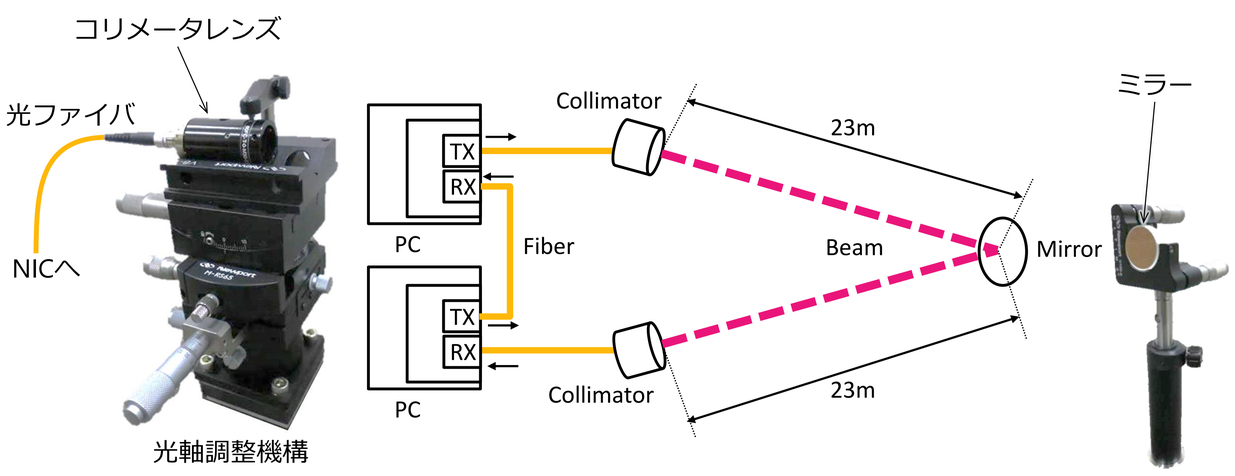
■光空間無線通信（FSO）





▲ ラック間ケーブルの 20% を FSO に置き換えると、平均ケーブル長が 27% 減る

- マシンラックの上にレンズを置き、ラック間を光ビームで接続
  - 10GBASE-LR／40GBASE-LR4 の赤外光を利用（λ=1310nm）
- 46m 離れたレンズ間でミラーを介した通信実験 [4]
  - **40Gbps** ワイヤレートで長時間安定した通信を確認
- ビームの向きを変えるだけでトポロジを変えられる
  - ランダムトポロジ化、パーティショニング、故障回復



発表論文

[4] Ikki Fujiwara, Michihiro Koibuchi, Tomoya Ozaki, Hiroki Matsutani, Henri Casanova: "Augmenting Low-latency HPC Network with Free-space Optical Links", The 21st IEEE International Symposium on High Performance Computer Architecture (HPCA 2015), Feb. 2015.

NII

連絡先：鯉渕道紘・藤原一毅 ／ 国立情報学研究所 アーキテクチャ科学研究系　准教授

TEL : 03-4212-2575　FAX : 03-4212-2120　Email : koibuchi@nii.ac.jp, ikki@nii.ac.jp

究極の電力効率を達成する

水没コンピュータ

In-Water Computers

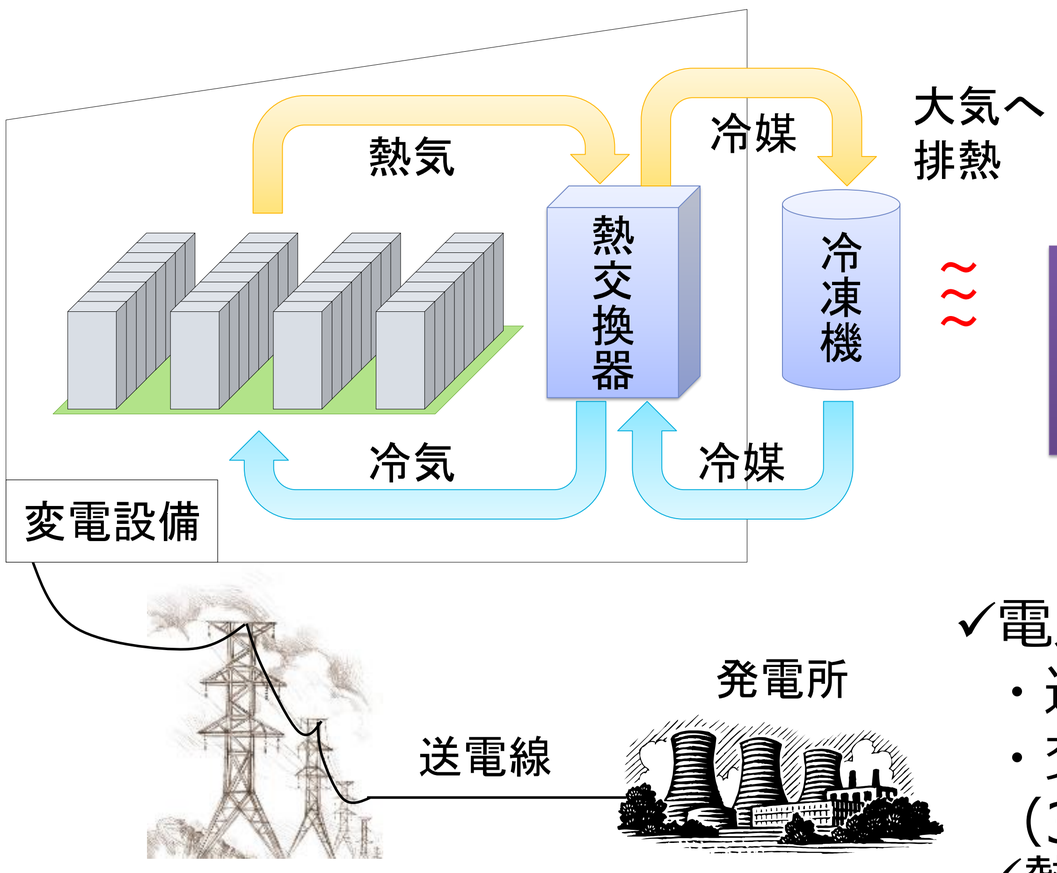
藤原一毅, 胡 曜, Truong Thao NGUYEN, 鯉渕道紘　I. Fujiwara, Y. Hu, T.T. Nguyen, M. Koibuchi

どんな研究？

もし，コンピュータが川や海などの水中で動作したら何ができるでしょうか？コンピュータは直接的に冷却され，水力発電のみならず，波力，潮力発電などの再生可能エネルギーを電力網の設置無しに直接的に利用することが可能となります．これは，電力の輸送と熱の輸送を極限まで抑えた「**無駄ゼロ**」コンピュータの実現を意味します．

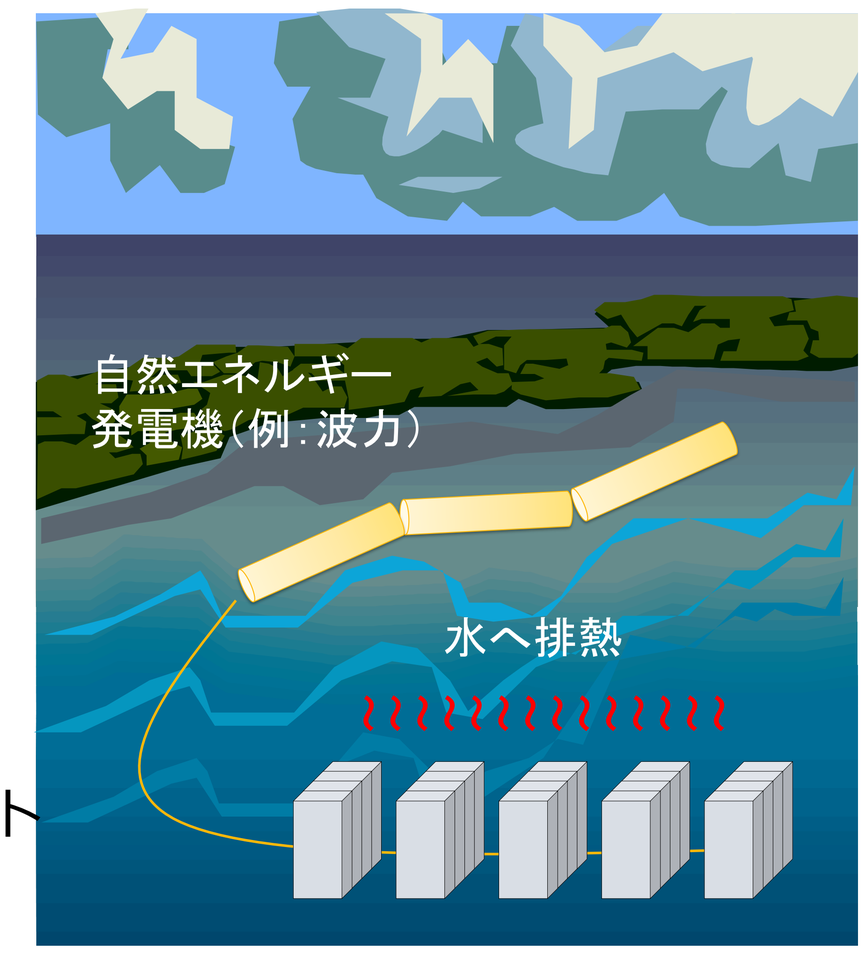
私たちは．水没コンピュータを実現するための要素技術，特に最適なコーティング技術の探究をしています。

状況設定



現在

✓電力を輸送しません
・送電ロス（5%）
・交直流変換ロスカット（30%）
✓熱を輸送しません
冷却設備の電力カット（30%～50%）



本研究ビジョン

どうやって水中でコンピュータを動作させるの？

私たちは今回，KISCO Ltd のdiX®コーティングを用いました．具体的には50–120 μmという薄いパリレンによる真空引きのコーティングです．これにより，コンピュータ内部が浸水することなく，水中に放熱を可能とします．データセンターのコンピュータの寿命は2年位なので，今後，2年間を目標にした耐久テストを行う予定です．

NII
連絡先：藤原一毅・鯉渕道紘 ／ 国立情報学研究所 アーキテクチャ科学研究系 准教授
TEL : 03-4212-2575　FAX : 03-4212-2120　Email : ikki@nii.ac.jp, koibuchi@nii.ac.jp

みんなの知恵で理想のスパコンを設計しよう

小直径グラフ探索コンペ “Graph Golf”

鯉渕道紘・藤原一毅・宇野毅明（国立情報学研究所）Michihiro Koibuchi, Ikki Fujiwara, Takeaki Uno (NII)
藤田聡・中野浩嗣（広島大学）Satoshi Fujita, Koji Nakano (Hiroshima University)

どんな研究？

理想のスーパーコンピュータをあなたも設計してみませんか？　私たちは、世界中の人々の力を借りて、**理論限界に迫る超低遅延ネットワーク・トポロジ**を探求しています。低遅延化にはランダムトポロジが有効なようですが、でたらめに作ったネットワークが本当に最善なのかどうか、実はまだ分かっていません。本研究では、究極の低遅延ネットワークの作り方を皆さんと一緒に解明していきます。

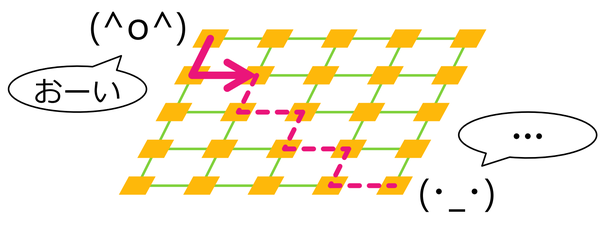
何がわかる？

スパコンのネットワークは、計算ノードを頂点、リンクを辺とする無向グラフです。頂点の数 n と、ひとつの頂点につながる辺の数 d は決まっています。あらゆる n, d の組合せに対し、最も低遅延なトポロジ、つまり**直径・平均距離が小さいグラフのカタログを作る**ことが私たちの目標です。この問題はグラフ理論家たちも挑戦してこなかったオープン問題で、本研究が世界初の取り組みとなります。

状況設定

■背景

- スーパーコンピュータのネットワークは大規模化と低遅延化が同時に求められている
- メッセージを端から端まで伝えるとき、途中で経由するノード数を最小限にしたい



▲従来型のネットワーク・トポロジ：途中で経由するノード数が多く、メッセージが届くのに時間がかかる

■問題定義：The Order/Degree Problem

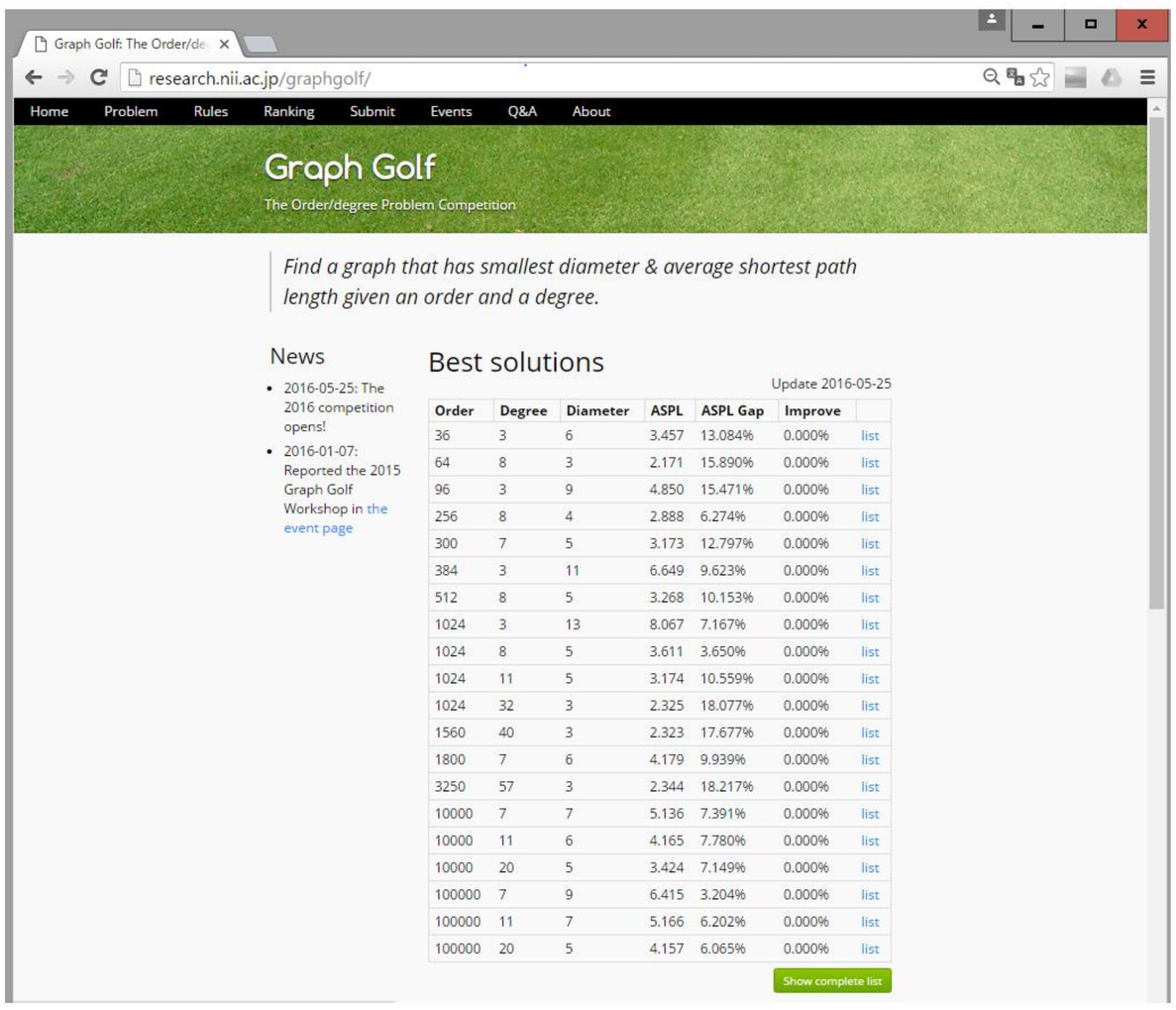
- 頂点数 n, 次数 d 以下の無向グラフを考える
  - ➢ 最も離れた２頂点間の最短ホップ数＝直径 k
  - ➢ すべての２頂点間の最短ホップ数の平均値＝平均距離 l
- 指定された n, d に対し、k, l が最も小さいグラフを求めよ
  - ➢ はじめに k を比較し、k が同じなら l を比較する

研究内容

■小直径グラフ探索コンペ “Graph Golf”

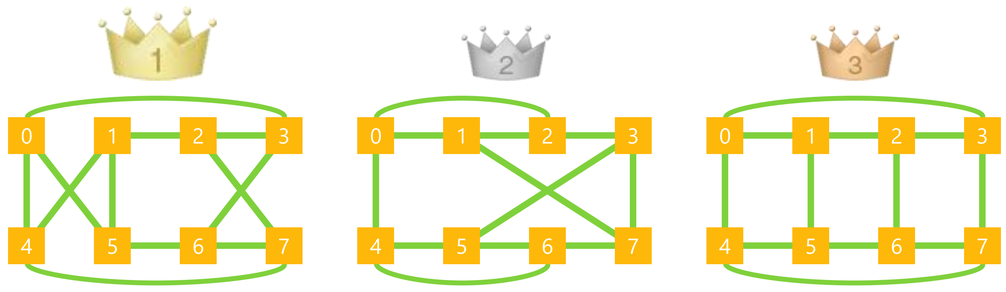
- 指定されたサイズ（頂点数・次数）で直径・平均距離の小さいグラフをホームページで募集。サイズを変えて毎年開催予定 [1]
- 優れたグラフの作者を国際会議 CANDAR で表彰

http://research.nii.ac.jp/graphgolf/



■解いてみよう

- 頂点数 n = 8, 次数 d = 3 で、直径・平均距離が最も小さいのはどんなグラフ？



グラフ			
直径 k	2	3	3
平均距離 l	1.571	1.643	1.714

距離表

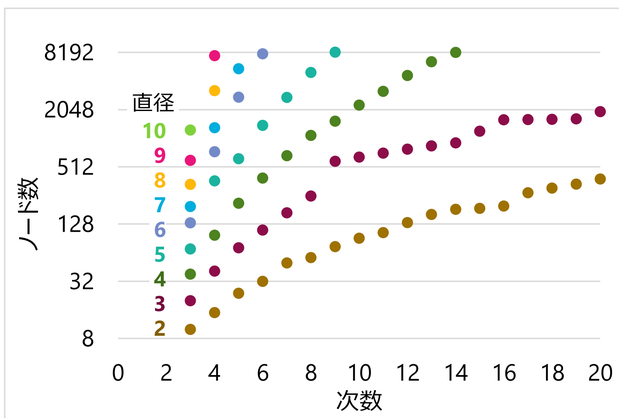
	0	1	2	3	4	5	6	7
0		2	2	1	1	1	2	2
1	2		1	2	1	1	2	2
2	2	1		1	2	2	2	1
3	1	2	1		2	2	1	2
4	1	1	2	2		2	2	1
5	1	1	2	2	2		1	2
6	2	2	2	1	2	1		1
7	2	2	1	2	1	2	1	

	0	1	2	3	4	5	6	7
0		1	1	2	1	2	2	2
1	1		1	2	2	3	2	1
2	1	1		1	2	2	3	2
3	2	2	1		2	1	2	1
4	1	2	2	2		1	1	2
5	2	3	2	1	1		1	2
6	2	2	3	2	1	1		1
7	2	1	2	1	2	2	1	

	0	1	2	3	4	5	6	7
0		1	2	1	1	2	3	2
1	1		1	2	2	1	2	3
2	2	1		1	3	2	1	2
3	1	2	1		2	3	2	1
4	1	2	3	2		1	2	1
5	2	1	2	3	1		1	2
6	3	2	1	2	2	1		1
7	2	3	2	1	1	2	1	

■関連問題：The Degree/Diameter Problem

- 指定された次数 d, 直径 k に対し、頂点数 n が最も大きいグラフを求める問題。求まる頂点数は飛び飛びの値になる
  - ➢ スパコンのノード数は予算や電力の制約で決まる。理論的に最適なノード数のマシンが造れるとは限らない



発表論文

[1] 藤原一毅, 藤田聡, 中野浩嗣, 井上武, 鯉渕道紘: “みんなでOrder/Degree問題を解いて究極の低遅延相互結合網をつくろう”, 2015年並列／分散／協調処理に関する『別府』サマー・ワークショップ (SWoPP2015), Aug. 2015.

NII
連絡先：鯉渕道紘・藤原一毅 ／ 国立情報学研究所 アーキテクチャ科学研究系　准教授
TEL : 03-4212-2575　FAX : 03-4212-2120　Email : koibuchi@nii.ac.jp, ikki@nii.ac.jp