Karlsruhe Institute of Technology

DIMACS

Center for Discrete Mathematics & Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

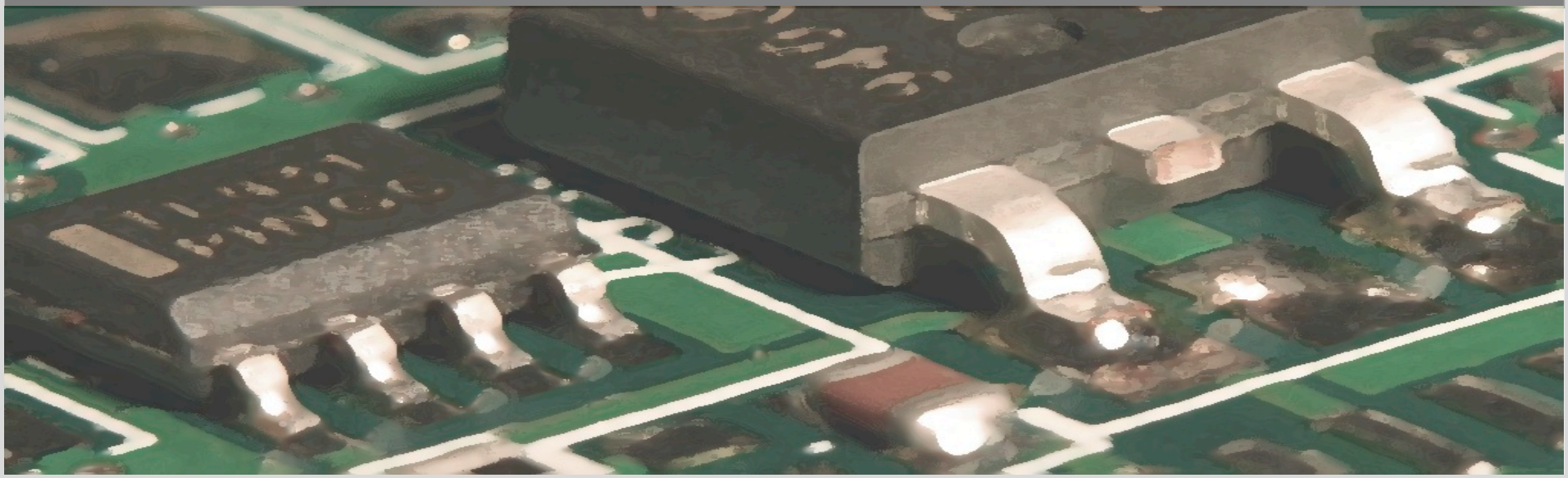
Georgia Tech College of Computing

10th DIMACS Implementation Challenge – Graph Partitioning and Graph Clustering – Results

David A. Bader
School of Computational Science and Engineering, Georgia Institute of Technology, USA

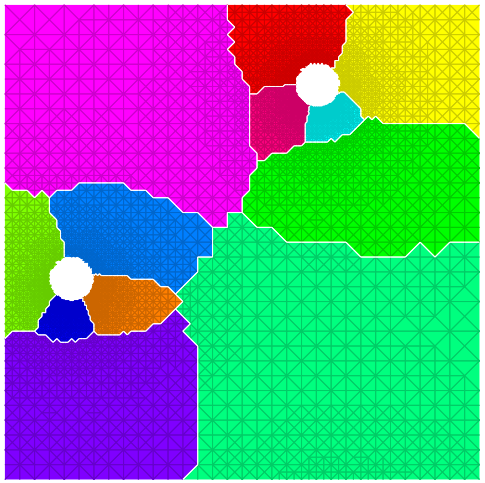
Henning Meyerhenke, Peter Sanders, Dorothea Wagner
with Christian Schulz and Andrea Schumm
Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Germany

Atlanta, Georgia · February 13/14, 2012

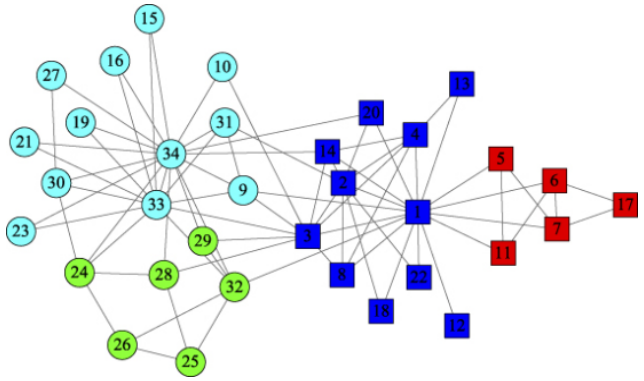


Challenge Types

- Graph Partitioning
  - Quality Challenge
    - Edge Cut
    - Communication Volume
  - Pareto Challenge
    - Quality
    - Work



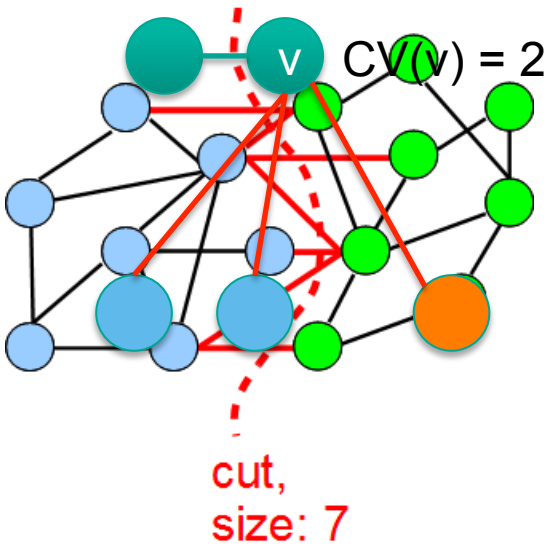
- Graph Clustering
  - Quality Challenge
    - Modularity
    - Mix
  - Pareto Challenge
    - Quality
    - Work



[Shen and Cheng, J. Stat. Mech. 2010]

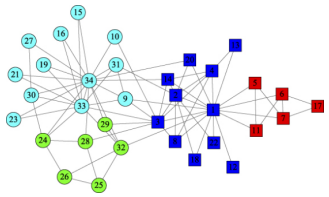
Graph Partitioning – Objectives

- **Edge Cut:** Size of cut
- **Communication Volume:** Number of boundary vertices, counted as often as they are adjacent to different parts



- **Rationale:**
  - EC well established
  - CV more accurate in some apps
- k power of 2, between 2 and 1024
- 3% imbalance
- 90 instances (18 graphs with 5 values of k each) of different sizes

Graph Clustering – Objectives



- **Modularity**

- **Mix:** Mixture of different objective functions
  - Performance
  - Average isolated inter-cluster conductance
  - Average isolated inter-cluster expansion
  - Minimum intra-cluster density

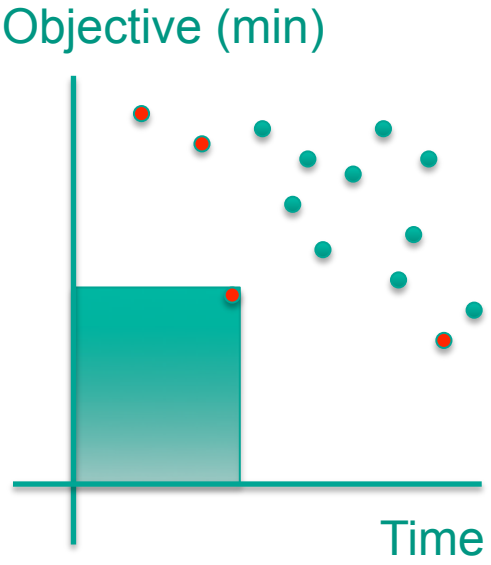
- **Rationale:**
  - Modularity popular, but has some flaws
  - Mix combines strengths (and weaknesses) of different measures

- 30 instances of different sizes

Quality Scoring

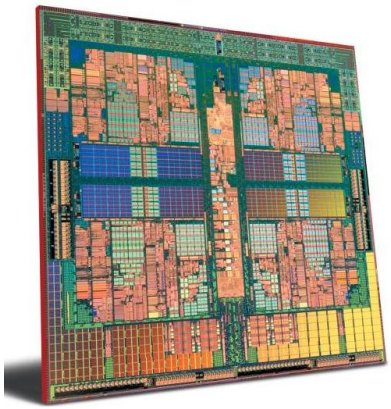
- Ranks given purely based on quality, F1 points for ranks: (10, 6, 4, 3, 2, 1)
- Each tuple (graph, k, objective) is an instance for which points are given
- **GP:** All points obtained are **accumulated** (EC + CV)
- **GC:** MO and MI are scored **separately**
  - Mix scoring rather complicated (look into the rules...)
- **The more points, the better!**

Pareto Challenge Scoring



- **Rationale:** Quality is important, but also your investment (work) to get it
- **Two dimensions:** Work and quality
- **Work:** (Wall-clock) Running time scaled by capability of your system
- **Special considerations:**
  - Multicore CPUs
  - Clusters
  - Accelerators

[techfresh.net]



[apcmag.com]

- Performance on **graph algorithms** important
- **Normalization factor:** (Number of cores)$^{0.9}$ / Benchmark timings (9th DIMACS Impl. Challenge)

CERTIFICATE OF EXCELLENCE

presented to

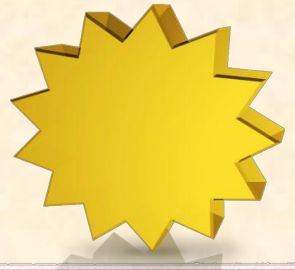
_____

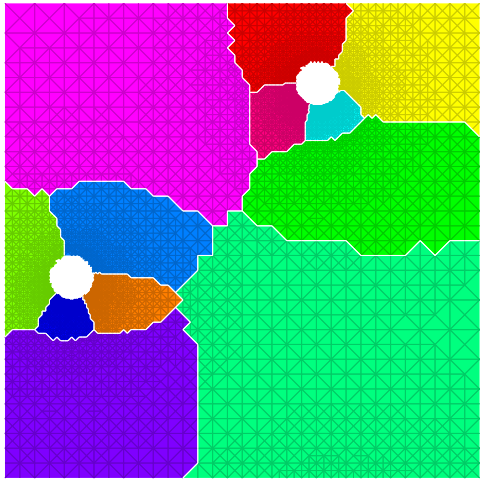
in recognition of your exceptional performance in the
10th DIMACS Implementation Challenge - Graph Partitioning and Graph Clustering,
held at Georgia Institute of Technology in Atlanta, Georgia,
February 13th through February 14th, 2012.

David A. Bader	Henning Meyerhenke	Peter Sanders	Dorothea Wagner
Georgia Institute of Technology	Karlsruhe Institute of Technology	Karlsruhe Institute of Technology	Karlsruhe Institute of Technology

DIMACS

Center for Discrete Mathematics & Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center





GRAPH PARTITIONING RESULTS

Results – Graph Partitioning

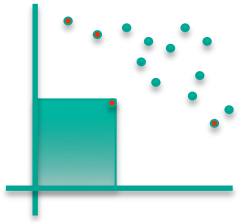
Quality Challenge (EC and CV combined)

- Three solvers:
  - UMPa (Catalyürek et al., OSU, ENS Lyon)
  - Mondriaan (Fagginger Auer and Bisseling, Utrecht U)
  - Karlsruhe Partitioners (KaFFPa etc.) (Sanders and Schulz, KIT)

- Formula 1 Points:
  - Karlsruhe Partitioners 1574
  - UMPa 1066
  - Mondriaan 616

- Number of Best Ranked Solutions:
  - Karlsruhe Partitioners 139
  - UMPa 25
  - Mondriaan 6

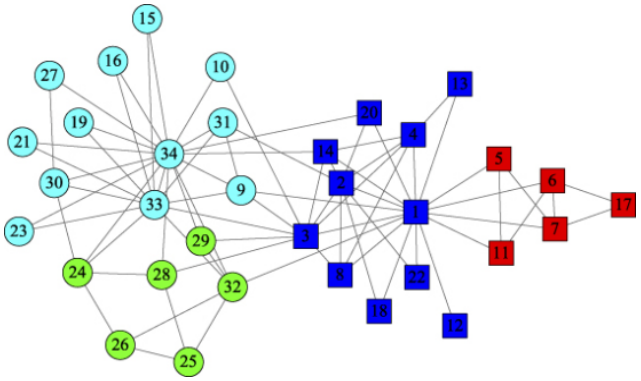
Results – Graph Partitioning
Pareto Challenge (EC and CV combined)



- Six solvers:
  - UMPa (Catalyürek et al., OSU, ENS Lyon)
  - Mondriaan (Fagginger Auer and Bisseling, Utrecht U)
  - KaFFPaFast, KaFFPaEco, KaFFPaE, KaFFPaStrong (Sanders and Schulz, KIT)

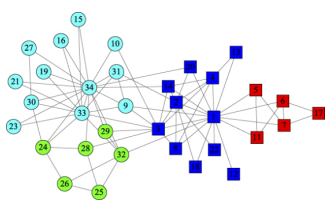
- Formula 1 Points:
  - KaFFPaFast 1680
  - KaFFPaEco 1305
  - KaFFPaE 1145
  - KaFFPaStrong 1106
  - UMPa Pareto 782
  - Mondriaan 462

- Number of Best Ranked Solutions:
  - KaFFPaFast 168
  - KaFFPaEco 122
  - KaFFPaE 109
  - KaFFPaStrong 100
  - UMPa Pareto 53
  - Mondriaan 27



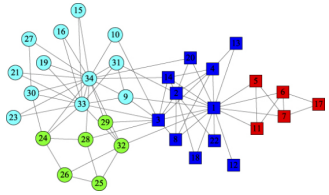
MODULARITY RESULTS

Results – Modularity
Quality Challenge: Solvers



- 15 solvers from 8 groups:
  - CGGCi_RG (Ovelgönne and Geyer-Schulz, KIT)
  - RG
  - RG+
  - ParMod (Catalyürek et al., OSU, ENS Lyon)
  - CLU_TBB (Fagginger Auer and Bisseling, Utrecht U)
  - CLU_CUDA
  - Scalable and Accurate... (Djidjev and Onus, LANL, Cankaya U)
  - community-el (Riedy et al., Georgia Tech, KIT)
  - community-el-xmt2
  - k-Community (Verma and Butenko, Texas A&M U)
  - k-Community++
  - VNS_quality (Aloise et al., UF Rio Grande et al.)
  - Stable Communities (Bhowmick and Srinivasan, UNO)
  - Stable Communities HD
  - Stable Communities BC

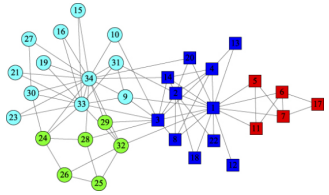
Results – Modularity
Quality Challenge

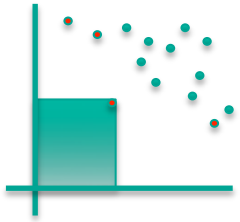


- Formula 1 Points:
  - CGGCi_RG 222
  - VNS_quality 196
  - ParMod 115
  - RG+ 100
  - RG 54
  - Scalable and Accurate... 34
  - community-el 16
  - community-el-xmt2 16
  - CLU_TBB 5
  - ... <=1
- # Rank 1:
  - CGGCi_RG 15
  - VNS_quality 11
  - ParMod 2
  - community-el 1
  - community-el-xmt2 1
  - ... 0

Results – Modularity

Pareto Challenge





- 10 (out of 15) solvers remain

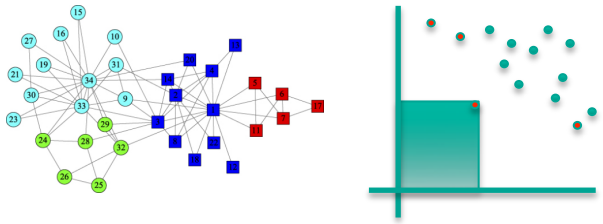
- Formula 1 Points:
  - RG 264
  - CGGCi_RG 256
  - RG+ 256
  - CLU_TBB 202
  - Scalable and Accurate... 117
  - VNS_pareto 116
  - community_el 59
  - CLU_CUDA 27
  - community-el-xmt2 17
  - ... <=5

- # Rank 1:
  - RG 26
  - CGGCi_RG 25
  - RG+ 25
  - CLU_TBB 19
  - VNS_pareto 9
  - Scalable and Accurate... 8
  - CLU_CUDA 2
  - community_el ?
  - ... 0

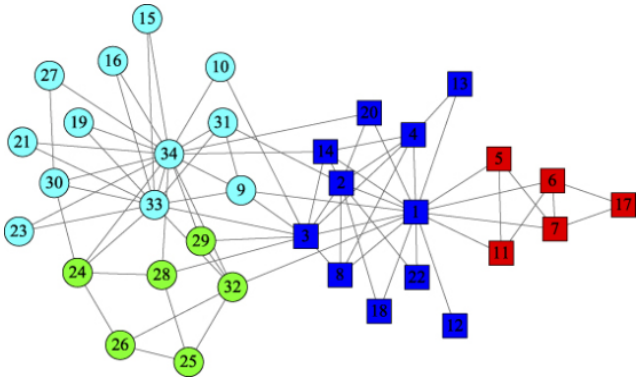
Results – Modularity

Pareto Challenge

without normalizing system performance

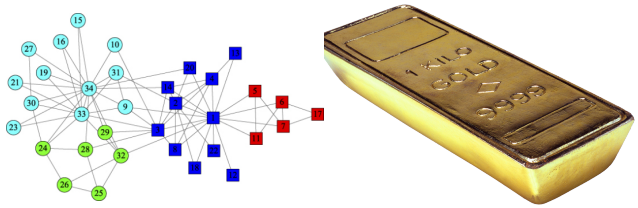


- Formula 1 Points:
  - RG+ 255
  - CLU_TBB 247
  - RG 246
  - CGGCi_RG 240
  - VNS_pareto 152
  - community-el 150
  - Scalable and Accurate... 111
  - CLU_CUDA 52
  - community-el-xmt2 18
  - ... 0
- Wall clock running time (**minutes**) for uk-2002 (n=18 520 486, m=261 787 258):
  - RG+ 228.5
  - CLU_TBB 0.5
  - RG 13.2
  - CGGCi_RG 7981.0
  - VNS_pareto
  - community-el 3.0
  - Scalable and Accurate...
  - CLU_CUDA
  - community-el-xmt2 12.9



MIX CHALLENGE RESULTS

Results – Mix
Quality Challenge



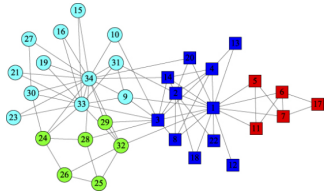
4 (out of 15) solvers remain:

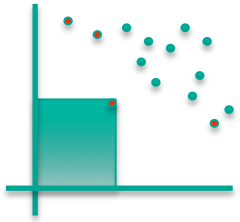
community-el (Riedy et al., Georgia Tech, KIT)
community-el-xmt2
k-Community (Verma and Butenko, Texas A&M U)
k-Community++

- Formula 1 Points:
  - community-el 155
  - community-el-xmt2 153
  - k-Community++ 146
  - k-Community 78
- # Rank 1:
  - community-el 10
  - community-el-xmt2 9
  - k-Community++ 9
  - k-Community 4

Results – Mix

Pareto Challenge



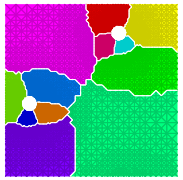


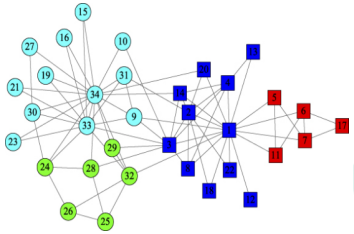
- 3 (out of 4) solvers remain

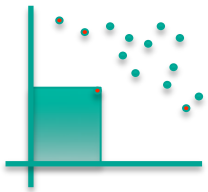
- Formula 1 Points:
  - community-el 182
  - k-Community++ 152
  - k-Community 110

- # Rank 1:
  - community-el 16
  - k-Community++ 10
  - k-Community 8

Challenge Summary







- Many submissions for modularity ☺
- Few for the two other categories ☹
- KaFFPa variants dominate graph partitioning (both competitors focus on hypergraphs)
- RG variants best for modularity, even in Pareto challenge
- Mixed challenge sees two entrants, community-el slightly better than k-Community++
- Pareto Challenge did not yield significantly different results
  - Likely due to Formula 1 scoring scheme

WORKSHOP SUMMARY

Workshop Summary

- Let‘s make this quick...
- Participation:
  - 19 contributed papers
  - GP: 3 groups
  - MO: 8 groups
  - MI: 2 groups
- Clustering has become the dominant topic
- Outreach:
  - Encyclopedia of SNA and Mining
  - ISMP 2012
- Please fill out your questionnaire
- Remove me from your SPAM list! ;-)

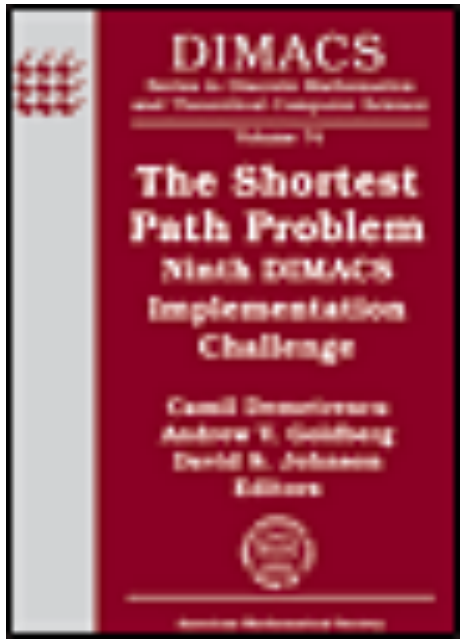
What happens next

- **Soon:** Several of the best workshop papers will be invited to write extended versions for an AMS-DIMACS book on the DIMACS Challenge Workshop

- Selection based on:
  - Scientific merit and novelty
  - Challenge results

- Tentative deadlines:
  - Submissions due in **March**
  - Review decision mailed in **May**
  - Final version due in **June**

- Details will be sent to authors of invited papers



Acknowledgments

People involved

- Advisory board:
  - Bruce Hendrickson, Sandia National Laboratories
  - David S. Johnson, AT&T Labs – Research
  - Timothy G. Mattson, Intel Corp, Comput. Software Laboratory
  - Chris Walshaw, University of Greenwich
- DIMACS:
  - Fred Roberts, Rebecca Wright
  - Nicole Clark, Linda Casals, …
- KIT:
  - Christian Schulz
  - Andrea Schumm
  - Robert Görke
- Georgia Tech:
  - Della Phinisee

T

H

A

N

K

S

!

Acknowledgments

Sponsors

- Provided travel grants, waived registration fees and other goodies!

Finally, thank you for your participation!!!