

VERİM

Veri Analitiđi
Arařtırma ve Uygulama Merkezi

Sabancı
Üniversitesi

YÜKSEK BAŞARIMLI VERİ ANALİTİĐİ VE MAKİNE ÖĐRENMESİ

KAMER KAYA

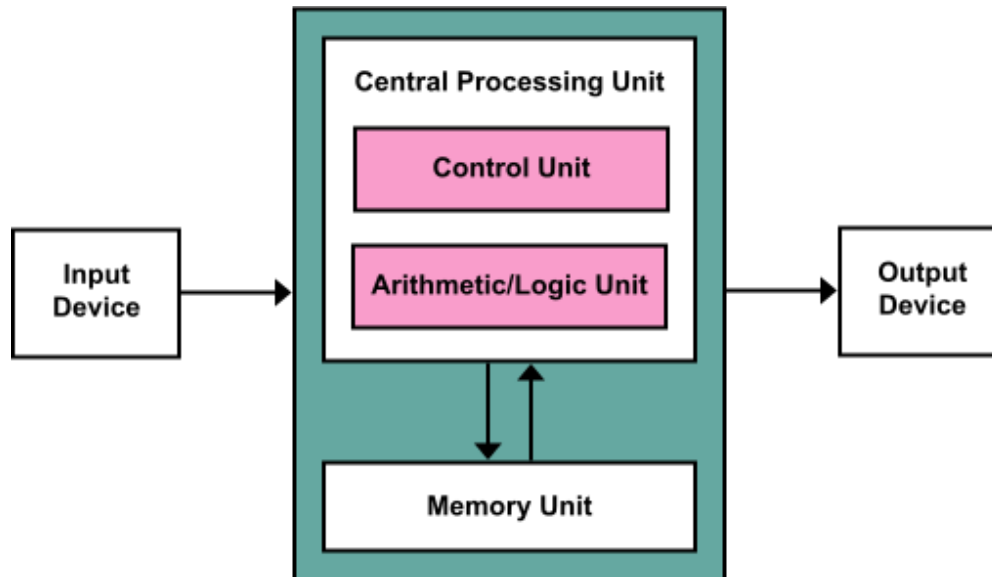
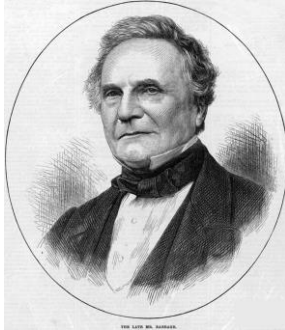
SABANCI ÜNİVERSİTESİ

MÜHENDİSLİK ve DOĐA BİLİMLERİ FAKÜLTESİ

Yüksek Başarım?

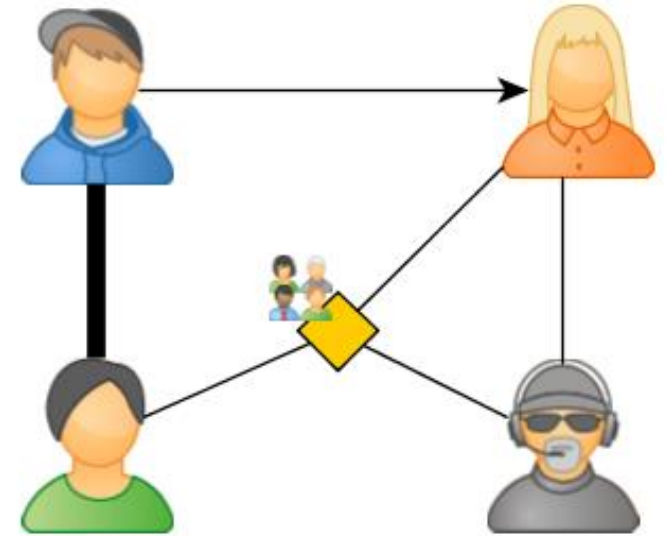
- 1. (isim) Elde edilen bir başarı
- 2. Herhangi bir olayı veya durumu başarma isteği ve gücü
- 3. Kişinin yapabileceği **en iyi** derece, performans
- 4. Herhangi bir eseri, oyunu, işi vb.ni ortaya koyarken gösterilen başarı, performans

Hesaplama

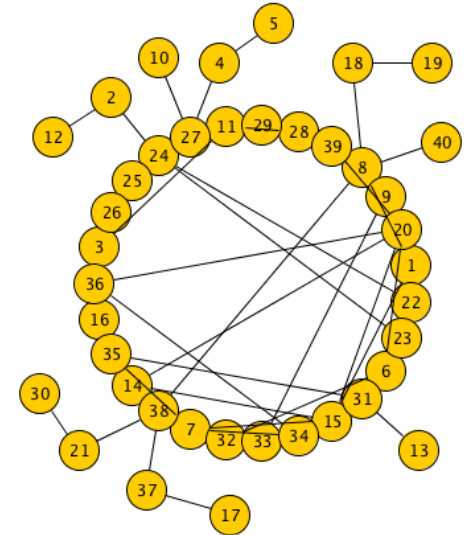
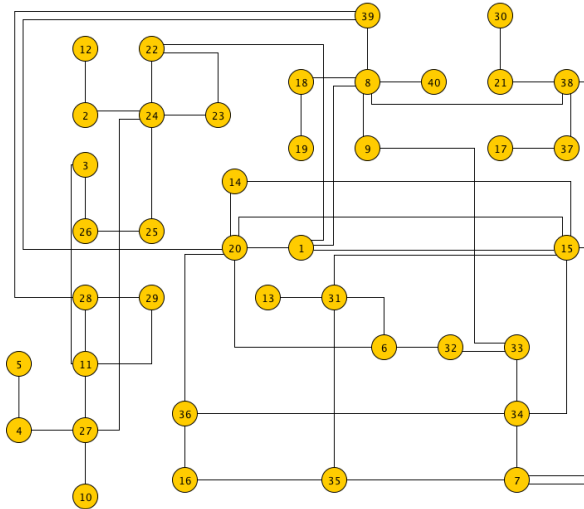
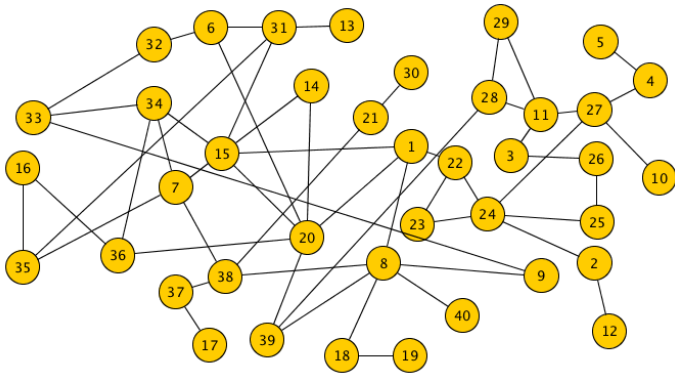


Çizge Veri Yapıları

- Çizgeler veri içerisindeki elemanların birbirleri ile olan ilişkilerini modellemek için kullanılan yapılardır.
- Farklı nokta (*vertex*) ve kenar (*edge*) türleri farklı tür ilişkileri modellemek için kullanılabilir.

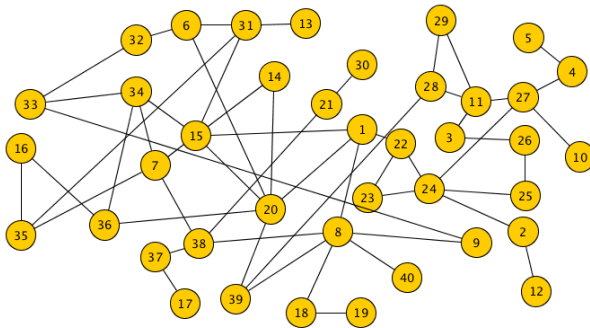


Çizge Verisi



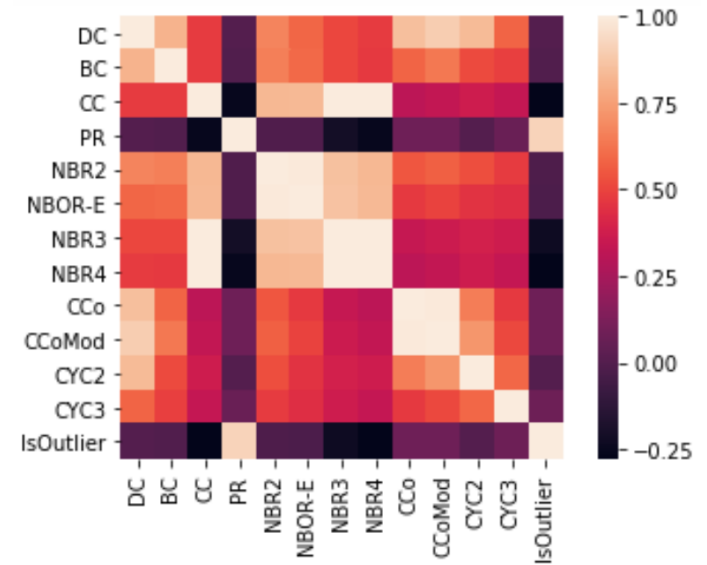
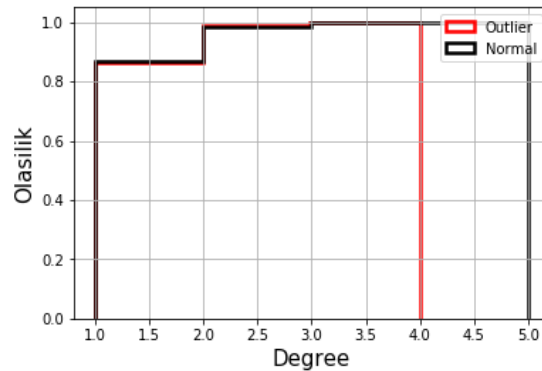
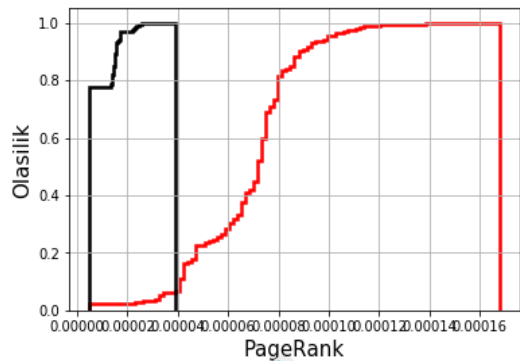
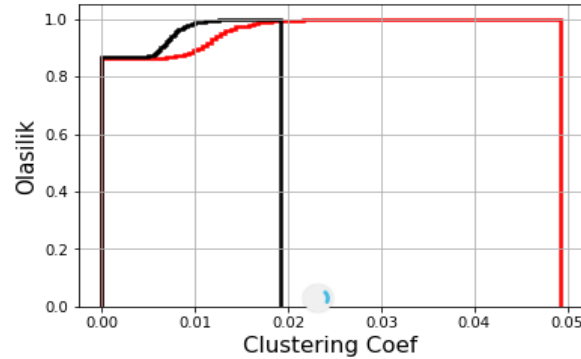
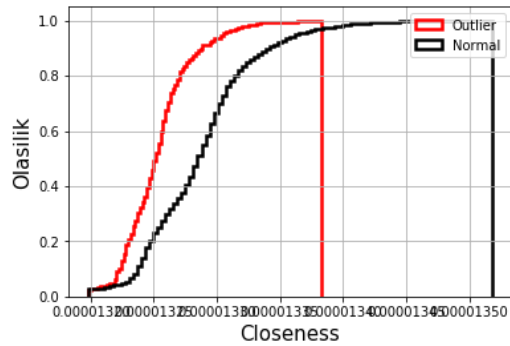
Noktaları Sınıflandırmak

- Bir sınıflandırma algoritmasında veri özellikleri önemli bir yere sahiptir.
- Çizge verisinde farklı sınıflardaki noktaların arasındaki ilişkiler farklı süreçler vasıtası ile oluşur.
 - Örneğin bir ağ üzerinde normal bir sunucunun iletişim şekli, zararlı bir sunucunun diğer sunucularla yaptığı iletişim şeklinden farklıdır.
 - Bu da noktaların, çizge üzerindeki yapısal özelliklerini farklılaştırır.

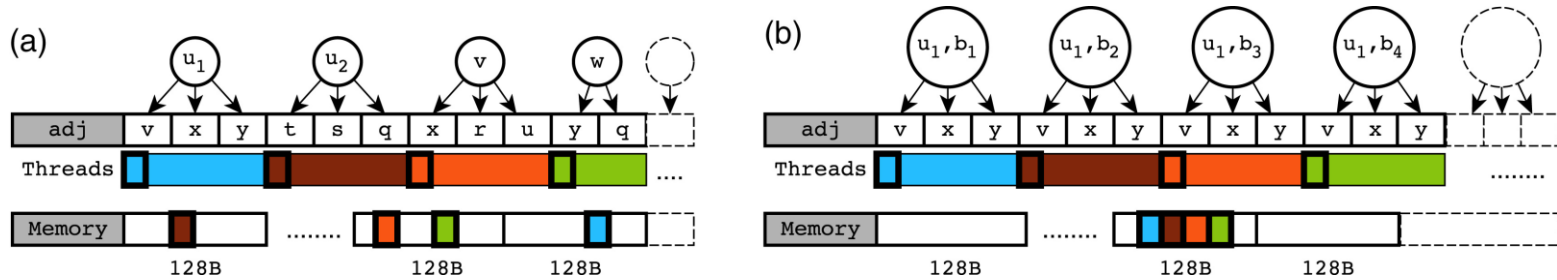


0.61	0.62	0.23	0.38
0.74	0.04	0.93	0.93
0.03	0.26	0.51	0.38
0.81	0.55	0.05	0.87
0.04	0.17	0.87	0.47
...
0.27	0.37	0.47	0.07
0.35	0.55	0.24	0.21
0.89	0.67	0.26	0.53

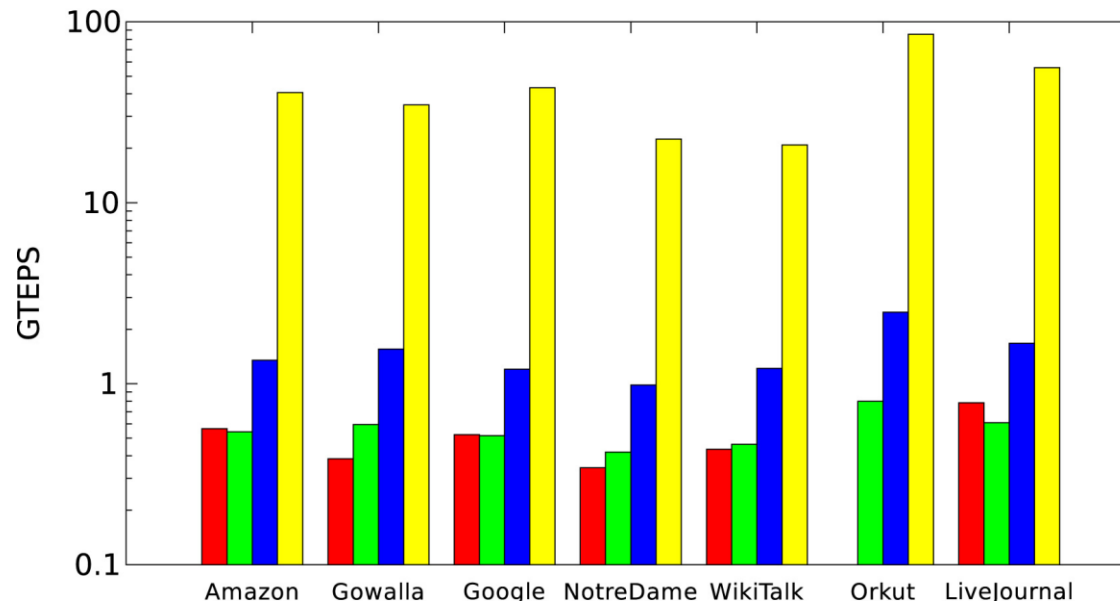
Noktaları Sınıflandırmak



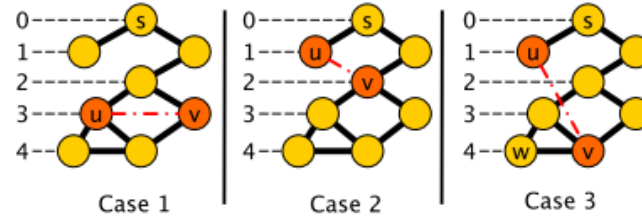
Yüksek Başarım



■ GPU-LinearBFS ■ GPU-VirCC-Multi
■ GPU-VirCC ■ GPU-SpMM



Yüksek Başarım



Graph			Time (in sec.)		
name	$ V $	$ E $	Org.	Best	Speedup
<i>hep-th</i>	8.3K	15.7K	1.41	0.05	29.4
<i>PGPgiantcompo</i>	10.6K	24.3K	4.96	0.04	111.2
<i>astro-ph</i>	16.7K	121.2K	14.56	0.36	40.5
<i>cond-mat-2005</i>	40.4K	175.6K	77.90	2.87	27.2
geometric mean					43.5
<i>soc-sign-epinions</i>	131K	711K	778	6.25	124.5
<i>loc-gowalla</i>	196K	950K	2,267	53.18	42.6
<i>web-NotreDame</i>	325K	1,090K	2,845	53.06	53.6
<i>amazon0601</i>	403K	2,443K	14,903	298	50.0
<i>web-Google</i>	875K	4,322K	65,306	824	79.2
<i>wiki-Talk</i>	2,394K	4,659K	175,450	922	190.1
<i>DBLP-coauthor</i>	1,236K	9,081K	115,919	251	460.8
geometric mean					99.8

Akan çizge verisi üzerinde yakınlık metriği hesaplaması için hızlanma oranları

Yüksek Başarımlı Veri Analitiđi

- OLAP (*online analytical proc.*) sorguları ve görevleri düşünülerek tasarlanmış, sütun tabanlı bir yapıya sahip,
- Verinin kendisini değil ama analiz sonuçlarını diferansiyel mahremiyete uygun şekilde gösterebilen,
- Bir ya da birden fazla CPU, FPGA ve GPU'yu aynı anda tek bir sorgu üzerinde, ya da farklı sorgular üzerinde akıllı iş dağıtım yöntemleriyle kullanabilen,
- Örneklem ve eskizleri yüksek başarımlı bir şekilde yaratarak doğruya yakın analiz sonuçlarının üretilmesi gibi farklı amaçlarla kullanabilecek,
- Kullanıcıların dinamik ve durağan sorgular üretebileceđi bir veritabanı yönetim sistemi.



Sabancı
Üniversitesi



Yürütücü:

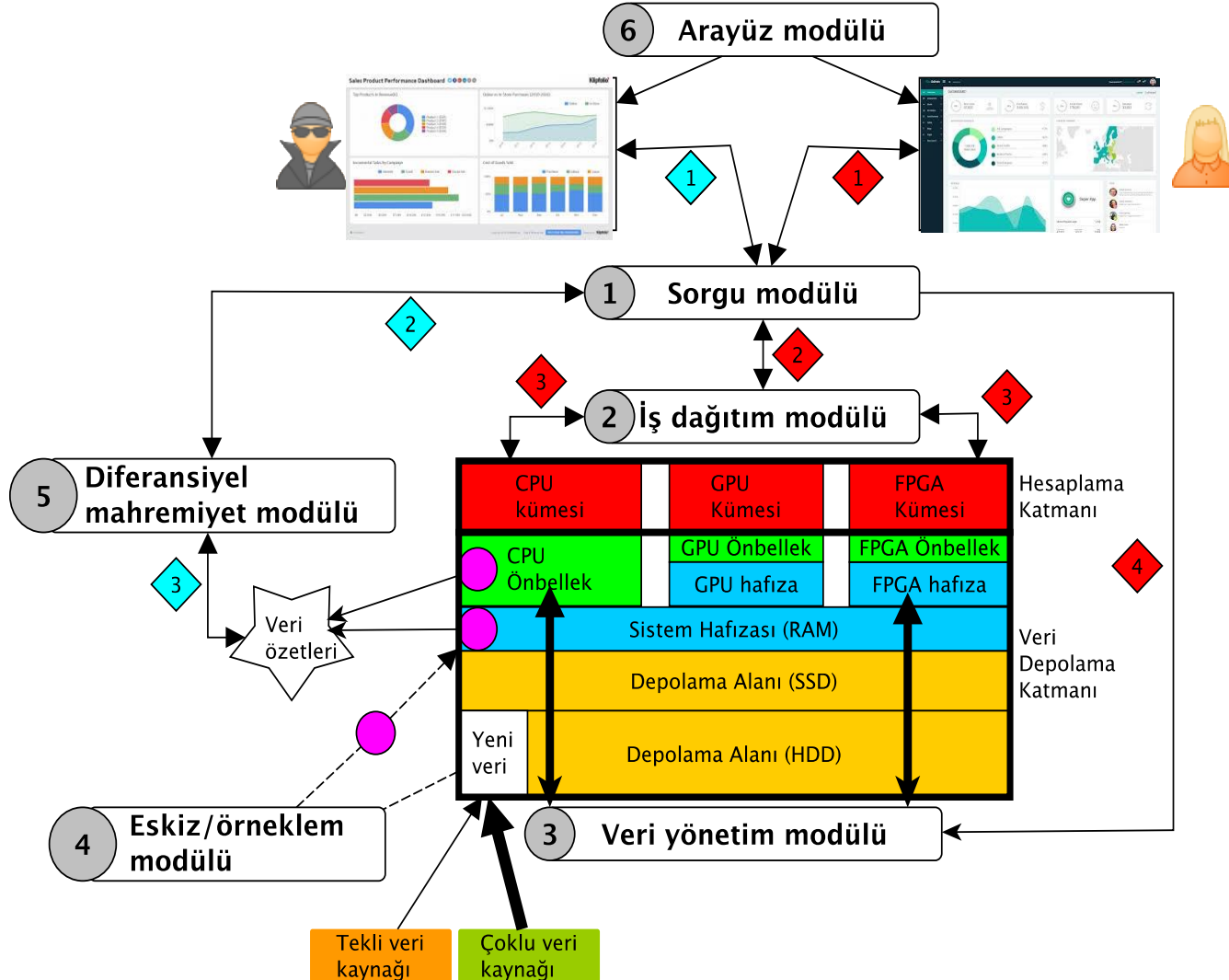
Dr. Sinan Yıldırım

Araştırmacılar:

Dr. Erdinç Öztürk,
Dr. Kamer Kaya

TÜBİTAK 1003

Yüksek Başarımlı Veri Analitiği



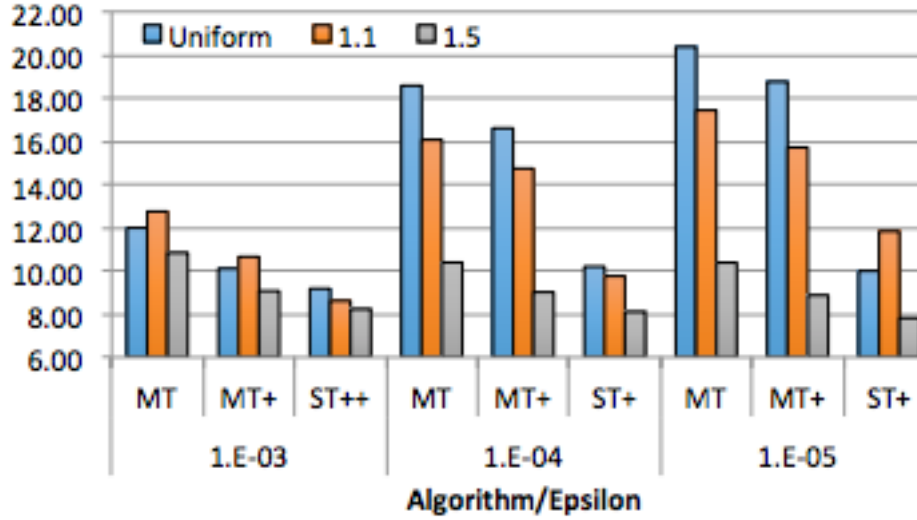
Sabancı Üniversitesi TÜBİTAK

Yürütücü:
Dr. Sinan Yıldırım

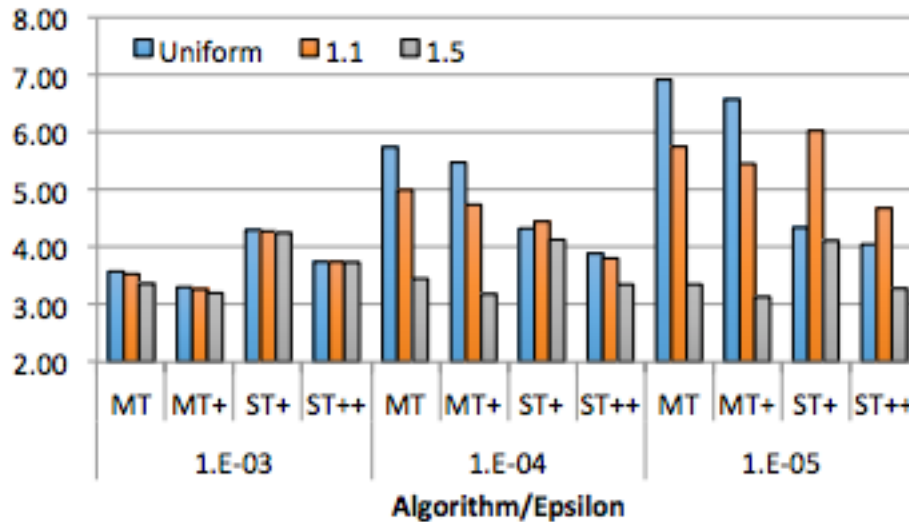
Araştırmacılar:
Dr. Erdiñç Öztürk,
Dr. Kamer Kaya

TÜBİTAK 1003

Yüksek Başarımlı Veri Analitiği



Raspberry Pi
3 Model B+



Odroid XU4

Sabancı
Üniversitesi



Yürütücü:
Dr. Sinan Yıldırım

Araştırmacılar:
Dr. Erdiñç Öztürk,
Dr. Kamer Kaya

TÜBİTAK 1003

VERİM

Veri Analitiği
Araştırma ve Uygulama Merkezi

DİNLEDİĞİNİZ İÇİN TEŞEKKÜRLER