

# SAPIENZA Università di Roma

Domanda di finanziamento per Progetti di Ricerca di Università  
Anno 2008 - prot. C26A08Y4NE

## 1. Dati Generali

### 1.1 Durata della ricerca

12 mesi

### 1.2 Responsabile della ricerca

**AUSIELLO**  
(cognome)

**Giorgio**  
(nome)

**Prof. Ordinario**  
(qualifica)

**30/07/1941**  
(data di nascita)

**ING. INFORMAZIONE, INFORMATICA e STATISTICA**  
(facoltà)

(dip/istit)

(indirizzo)

**06/77274005**  
(telefono)

**06/77274002**  
(fax)

**ausiello@dis.uniroma1.it**  
(e-mail)

### 1.4 Titolo della ricerca

*Algoritmi e strutture informative per grandi moli di dati.*

## 2. Informazione sull'attività di ricerca

### Area su cui insiste il progetto

09 - Ingegneria industriale e dell'informazione

### 2.1 Parole chiave

1. ALGORITMI
2. STRUTTURE DI DATI
3. DATA STREAM
4. GRAFI
5. WORLD WIDE WEB

## 2.2 Ambito della ricerca

Interfacoltà

## 2.3 Tipologia

Continuazione

## 2.4 Componenti il gruppo di ricerca (escluso il responsabile)

### Personale docente dell'Università

n°	Cognome	Nome	Qualifica	Facoltà	Dipartimento
1.	BECCHETTI	Luca	RU	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA
2.	D'AMORE	Fabrizio	PA	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA
3.	DEMETRESCU	Camil	RU	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA
4.	FERRARO PETRILLO	Umberto	RU	SCIENZE STATISTICHE	DIP. STATISTICA,PROBABILITA' E STATISTICHE APPL.
5.	FRANCIOSA	Paolo Giulio	PA	SCIENZE STATISTICHE	DIP. STATISTICA,PROBABILITA' E STATISTICHE APPL.
6.	LEONARDI	Stefano	PO	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA
7.	MARCHETTI SPACCAMELA	Alberto	PO	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA
8.	VITALETTI	Andrea	RU	INGEGNERIA	DIP. INFORMATICA E SISTEMISTICA

### Altro personale dell'Università "Sapienza" di Roma

n°	Cognome	Nome	Qualifica	Facoltà	Dipartimento	Note
1.	RIBICHINI	ANDREA	Professore a Contratto			
2.	DONATO	DEBORA	Professore a Contratto			
3.	LAURA	LUIGI	Professore a Contratto			

### Personale di altre Università/Istituzioni

n°	Cognome	Nome	Qualifica	Università/Istituzione	Dipartimento	Note
1.	BONIFACI	VINCENZO	Assegnista di ric.	Università' de L'Aquila		

## 2.5 Inquadramento della ricerca proposta (in ambito nazionale ed internazionale)

Negli ultimi dieci anni gli sviluppi nelle applicazioni informatiche hanno posto sempre maggiore enfasi sulla gestione di grandi moli di dati, come collezioni di documenti di grandi dimensioni, grafi e matrici di grande taglia. La necessità di gestire grandi moli di dati sorge ad esempio in bioinformatica, in meteorologia, nella soluzione di problemi computazionali difficili, e nella visualizzazione ed esplorazione di grandi reti informatiche (come Internet) o sociali. Per varie ragioni, progettare algoritmi efficienti per grandi moli di dati può essere un compito molto arduo. Tipicamente, i dati sono memorizzati in memoria esterna e questo rende necessario l'impiego di tecniche algoritmiche per la loro gestione in modo da minimizzare il numero di accessi a disco. Trattare con dati di grandi dimensioni solleva inoltre numerose questioni algoritmiche connesse con l'uso efficiente delle memorie gerarchiche di cui sono dotati i moderni sistemi di calcolo. Questo scenario complesso rende inoltre necessario l'uso di nuove tecniche di ingegnerizzazione e visualizzazione. Si noti che talvolta le dimensioni dei dati generati dalle applicazioni possono essere talmente grandi da renderne difficoltosa la memorizzazione perfino in memoria esterna, richiedendone quindi un'analisi di tipo streaming. Questo progetto di ricerca ha i seguenti obiettivi:

- (1) scoprire nuove tecniche e nuove metodologie algoritmiche per elaborare strutture informative di grandi dimensioni;
- (2) identificare e risolvere problemi di natura algoritmica, cruciali in importanti applicazioni in cui si richiede la gestione di grandi moli di dati;
- (3) contribuire ad un rapido trasferimento di tecnologie algoritmiche evolute mediante sperimentazione ed ingegnerizzazione di codici di algoritmi efficienti.

In particolare la ricerca si articola in due direzioni complementari:

- MODELLI COMPUTAZIONALI E METODOLOGIE ALGORITMICHE
- ALGORITMI PER GRAFI DI GRANDI DIMENSIONI

Con riferimento al primo obiettivo studieremo le proprietà di vari modelli computazionali utilizzati per gestire strutture informative di grandi dimensioni (come i modelli per memorie gerarchiche e il modello 'data stream') e le tecniche algoritmiche più idonee per i diversi modelli. Svilupperemo inoltre strumenti per l'ingegnerizzazione e visualizzazione di algoritmi su dati di grandi dimensioni.

Rispetto al secondo obiettivo studieremo campi applicativi reali in cui la capacità di elaborare efficientemente strutture informative di grandi dimensioni gioca un ruolo cruciale per ridurre i costi di elaborazione ed accrescere efficienza ed affidabilità. In particolare ci concentreremo sul progetto di algoritmi per grafi e reti di grandi dimensioni. In questo ambito un ruolo fondamentale avrà chiaramente lo studio di algoritmi per analizzare le proprietà del web o di altre reti su larga scala (ad esempio reti P2P).

## 2.6 Sintesi del programma di ricerca e descrizione dei compiti dei singoli partecipanti

In questa sezione descriviamo gli aspetti più rilevanti del programma di ricerca proposto, discutendo per ciascuna tematica le problematiche generali affrontate, il loro collocamento rispetto alla letteratura di riferimento, e gli obiettivi e i risultati parziali attesi. In questo contesto sono descritti gli obiettivi raggiunti durante il primo anno del progetto; nel secondo anno, i ricercatori impegnati nel progetto continueranno a cooperare per il conseguimento degli obiettivi, partecipando attivamente sia alla prima linea di ricerca di tipo più metodologico, che a quella finalizzata allo studio di grafi e reti di grandi dimensioni.

## MODELLI COMPUTAZIONALI E METODOLOGIE ALGORITMICHE

In questo progetto studiamo vari modelli computazionali utilizzati per gestire strutture informative di grandi dimensioni come quelli per gerarchie di memoria e il modello 'data stream'. In particolare studiamo come le tecniche di progetto di algoritmi e gli aspetti tecnologici dei sistemi con memoria complessa interagiscono e si adattano l'un l'altro. Inoltre ci dedichiamo anche allo studio di tecniche algoritmiche generali, come metodi di riduzione di algoritmi tra modelli computazionali diversi e per la visualizzazione di grafi di dimensioni molto grandi. Un altro scopo è quello di sviluppare strumenti e tecniche per ingegnerizzare algoritmi per grandi moli di dati.

### Modelli di calcolo per grandi moli di dati

Un primo aspetto che intendiamo studiare è il trattamento di dati che vengono generati sotto forma di stream di dati. Questo scenario si presenta in molte applicazioni scientifiche che generano enormi quantità di dati soggetti a rapidi cambiamenti, che devono essere continuamente tenuti sotto controllo per effettuare analisi ed individuare correlazioni, eventi rari, frodi, intrusioni, attività inusuali o anomale. Esempi rilevanti includono il controllo del traffico di rete, aste online, transazioni, informazioni sulle chiamate telefoniche, operazioni bancarie, eventi atmosferici ed astronomici.

Gli stream di dati devono essere processati in una o poche passate sequenziali usando una quantità limitata di memoria di lavoro.

Nonostante le pesanti restrizioni sulle risorse di spazio e di tempo imposte da tale modello, molti problemi di sintesi dei dati e calcolo di statistiche sono stati affrontati con successo, come frequency moments, quantili, istogrammi e wavelet. La maggior parte di tali problemi possono essere approssimati con spazio e passate polilogaritmiche nella dimensione dello stream [M05]. Più recentemente, l'interesse si è spostato sulla soluzione di problemi su grafi [FKMSZ04]. Molti classici problemi su grafi sembrano essere difficili nel contesto di streaming e richiedono tipicamente  $\Omega(n/s)$  passate, dove  $n$  è il numero di vertici ed  $s$  è la dimensione della memoria di lavoro. Problemi come accoppiamento, bipartitezza, connettività, minimo albero di copertura,  $t$ -spanner, possono essere tutti risolti in  $O(1)$  passate con spazio  $O(n \log n)$ .

Motivati da ragioni tecnologiche, vari ricercatori hanno recentemente proposto modelli meno restrittivi, in cui gli algoritmi possono scrivere stream intermedi o persino usare primitive di ordinamento sugli stream, progettando algoritmi efficienti in memoria esterna basati sul paradigma di streaming [ADRR04, DFR06]. Data l'alta velocità di accesso sequenziale nei moderni dischi, algoritmi di streaming possono essere usati efficacemente per processare file di grandi dimensioni, affinando le tecniche di soluzione di problemi in memoria esterna [M05].

Un ulteriore modello per grandi moli di dati considerato è il modello cache-oblivious [K03], in cui gli algoritmi si adattano automaticamente alla particolare piattaforma hardware senza doverne conoscere i dettagli e traendo vantaggio in particolare dalla presenza di gerarchie di memoria caratterizzate da diverse velocità di accesso.

### Obiettivi raggiunti nel primo anno:

Abbiamo studiato come trasformare algoritmi paralleli in algoritmi di streaming per vari problemi di calcolo, fornendo nuove tecniche di gestione di grandi moli di dati in memoria esterna. Una lista parziale dei problemi studiati include ordinamento, connettività, minimo albero ricoprente, componenti biconnesse e massimo insieme indipendente.

A questo scopo, sono state progettate tecniche generali di simulazione di algoritmi PRAM nel modello W-Stream proposto in [R03].

Dal punto di vista del modello cache-oblivious, è stata messa a punto un'infrastruttura per sperimentare algoritmi in grado di adattarsi automaticamente ai parametri hardware che definiscono ad esempio le dimensioni dei blocchi trasferiti tra livelli in una memoria gerarchica.

### Risultati del primo anno:

- [DEMR07] Tecniche di simulazione PRAM nel modello W-Stream
- [ADFIR08] Nuovi algoritmi di streaming basati sulle tecniche di simulazione proposte e in altri modelli classici di streaming
- [ADP07] Piattaforma di sperimentazione di algoritmi cache-oblivious

### Obiettivi del secondo anno:

In questa fase considereremo questioni pratiche connesse con l'implementazione e analisi sperimentale degli algoritmi sviluppati nel primo anno. In particolare, effettueremo un'indagine sperimentale delle prestazioni di algoritmi W-Stream e di precedenti algoritmi in memoria esterna per problemi fondamentali come connettività e cammini minimi in grafi.

I nostri codici saranno basati su librerie efficienti per processare grandi moli di dati come la STXXL. I nostri programmi saranno rilasciati nel pubblico dominio come collezione di algoritmi efficienti di streaming.

Contiamo infine di analizzare sperimentalmente gli algoritmi di tipo cache-oblivious sviluppati nel contesto di questo progetto, utilizzando la piattaforma sviluppata durante il primo anno.

### Risultati parziali attesi dal secondo anno:

- Confronto sperimentale di algoritmi W-stream e algoritmi noti su memoria esterna per connettività e cammini minimi.
- Collezione software di algoritmi efficienti di streaming su grafi.
- Analisi sperimentale di algoritmi cache oblivious nella piattaforma sviluppata nel primo anno.

### Metodologie e strumenti di visualizzazione di algoritmi per grandi moli di dati

Negli ultimi anni la quantità di dati elaborati elettronicamente è cresciuta a tal punto da rendere necessario l'impiego di tecniche completamente nuove [Vi0]. In questo contesto, sviluppare software algoritmico robusto ed efficiente è un compito spesso delicato e difficoltoso, data la necessità di accedere a dati memorizzati su diversi livelli di un sistema di memoria complesso. Strumenti visuali, ampiamenti assenti negli ambienti di sviluppo convenzionali, possono essere di grande aiuto nell'analisi, messa a punto e debugging di programmi in questo scenario.

Sebbene molti sistemi software e librerie siano stati proposti nell'ultima decade per visualizzare diversi aspetti di un programma algoritmico [DFIN02], trattare con software per grandi moli di dati rimane un problema tutt'ora aperto. In particolare, produrre visualizzazioni integrate di dati visualizzati a differenti livelli di un sistema di memoria gerarchico rimane un compito difficile.

### Obiettivi raggiunti nel primo anno:

Sono stati sviluppati strumenti di visualizzazione integrata di dati memorizzati a diversi livelli di un sistema di memoria. In particolare, ci si è concentrati su algoritmi per grafi e reti, sia in contesti di memoria esterna che di streaming, e usando tecniche di disegno di grafi.

### Risultati ottenuti nel primo anno:

- [BDFL06] Progetto e implementazione di tool per la visualizzazione e l'ingegnerizzazione di codice algoritmico per grandi moli di dati

### Obiettivi del secondo anno:

I tool di visualizzazione sviluppati nel primo anno verranno valutati e testati su problemi specifici studiati in questo progetto. Ad esempio, ci proponiamo di valutare l'efficacia dei tool proposti per analizzare alcune delle implementazioni di algoritmi di streaming.

### Risultati parziali attesi del secondo anno:

- Valutazione dei tool di visualizzazione di algoritmi sviluppati nel primo anno su specifici problemi su grafi come connettività e cammini minimi.

## ALGORITMI PER GRAFI DI GRANDI DIMENSIONI

Un importante dominio applicativo considerato in questo progetto è la gestione di reti. Il nostro obiettivo è quello di studiare tecniche innovative per la gestione di grafi di grandi dimensioni, anche a fronte di modifiche del grafo, concentrandoci sia su questioni metodologiche che di ingegnerizzazione di software algoritmico per applicazioni di rete.

In applicazioni come l'instradamento di pacchetti in Internet, la gestione efficiente del traffico si basa sul mantenimento di informazioni derivate dalla struttura della rete stessa. Un esempio rilevante è il mantenimento di cammini minimi tra router, tipicamente usato da protocolli standard di instradamento come OSPF [M99] per evitare congestioni del traffico.

Un altro problema che intendiamo studiare sia da un punto di vista analitico che da un punto di vista sperimentale è il calcolo di sottografi sparsi (spanner) che preservino le distanze in un grafo in modo approssimato. Studieremo inoltre la possibilità di estrarre informazioni approssimate (sketches) da uno stream di dati rappresentante un grafo o l'istanza di un problema combinatorio (es. set cover). Infine, ci proponiamo di sviluppare algoritmi su grafi nel modello cache-oblivious [K03], in cui come già osservato gli algoritmi sono ottimizzati per trarre vantaggio dalla presenza di gerarchie di memoria caratterizzate da diverse velocità di accesso.

Obiettivi raggiunti nel primo anno:

Un primo obiettivo era quello di progettare algoritmi dinamici per cammini minimi in grado di ottimizzare le risorse di memoria, con particolare riferimento a memoria esterna e uso delle cache. Particolare enfasi è stata data a soluzioni algoritmiche per problemi su grafi di grandi dimensioni di interesse industriale come il mantenimento dinamico di tabelle di instradamento, critico per le aziende di telecomunicazioni.

In particolare, sono stati realizzati algoritmi centralizzati e distribuiti per la raccolta di pacchetti in reti wireless. Sono stati inoltre sviluppati algoritmi efficienti nei modelli data stream e memoria esterna per il calcolo di spanner di grafi; uno spanner può essere sfruttato per risolvere problemi di cammini minimi su grafi di grandi dimensioni.

Sono stati studiati metodi di analisi del grafo Web per la individuazione di Web spam.

Sono stati progettati infine algoritmi cache-oblivious efficienti per alcuni problemi fondamentali su grafi

Risultati ottenuti nel primo anno:

- [AFI08, ADFIR08] Progetto e sperimentazione di algoritmi in data streaming per il calcolo di spanners
- [BBCG08] Algoritmi in semi-streaming per il calcolo del numero di triangoli incidenti su ciascun vertice in un grafo
- [BCDBL08] Algoritmi basati sulla struttura dei link del Web per la individuazione di Web Spam.
- [BKMS08a, BKMS08b, BKMS08c] algoritmi centralizzati e distribuiti per la raccolta di pacchetti in reti wireless
- [DFR08] Tecniche algoritmiche per grafi dinamici in memoria esterna
- [ALZ07] Nuovi algoritmi cache-oblivious per alcuni problemi fondamentali su grafi

Obiettivi del secondo anno:

Ci proponiamo di effettuare una approfondita valutazione sperimentale di algoritmi dinamici per cammini minimi su grafi di grandi dimensioni, distribuendo il codice prodotto nel pubblico dominio e usando librerie software standard.

Osserviamo inoltre che quasi tutti gli algoritmi teoricamente veloci sulle gerarchie di memoria sono difficili da implementare.

Tuttavia, molte tecniche algoritmiche sviluppate in questo campo possono essere rilevanti anche nella pratica.

Nel contesto degli algoritmi per grafi lavoreremo con l'obiettivo di rendere (parzialmente) dinamici gli approcci usati negli algoritmi statici. Parallelamente sarà portata avanti l'attività di ricerca, per raffinare e semplificare gli algoritmi sviluppati nel primo anno.

Risultati parziali attesi del secondo anno:

- Implementazione e ingegnerizzazione degli algoritmi dinamici progettati nel primo anno
- Validazione sperimentale di algoritmi esterni su grafi

Riferimenti bibliografici

- [ADRR04] G. Aggarwal, M. Datar, S. Rajagopalan, M. Ruhl. On the streaming model augmented with a sorting primitive. Proc. IEEE FOCS, 540-549, 2004.
- [DFIN02] C. Demetrescu, I. Finocchi, G.F. Italiano, and S. Naeher. Visualization in Algorithm Engineering: Tools and Techniques. In "Experim. Algorithmics, From Algorithm Design to Robust and Efficient Softw.", LNCS 2547, 2002.
- [DFR06] C. Demetrescu, I. Finocchi, and A. Ribichini. Trading off space for passes in graph streaming problems. Proc. SODA, 714-723, 2006.
- [FKMSZ04] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, J. Zhang. On graph problems in a semi-streaming model. Proc. ICALP, 531-543, 2004.
- [K03] P. Kumar. Cache-Oblivious Algorithms. Algorithms for Memory Hierarchies, LNCS 2625, 193-212, 2003.
- [M05] S. Muthukrishnan. Data streams: algorithms and applications. <http://athos.rutgers.edu/~muthu>.
- [M99] J. Moy, "OSPF: Anatomy of an Internet Routing Protocol", Addison-Wesley, 1999.
- [R03] J. Ruhl. Efficient Algorithms for New Computational Models. PhD thesis, MIT, 2003.
- [Vi01] J.S. Vitter. External Memory Algorithms and Data Structures: Dealing with MASSIVE DATA, ACM Comp. Surv., 33(2):209-271, 2001.

### **3. Elenco delle migliori pubblicazioni negli ultimi 5 anni**

#### **A) Pubblicazioni su riviste scientifiche**

1. AUSIELLO G.; M. DEMANGE; L. LAURA; V. PASCHOS (2005). Algorithms for the On Line Quota Traveling Salesman Problem. INFORMATION PROCESSING LETTERS, vol. 92, 2; p. 89-94, ISSN: 0020-0190
2. DUBASHI D, HAGGSTROM O, ORECCHIA L, PANCONESI A, PETRIOLI C, VITALETTI A. (2007). Localized techniques for broadcasting in wireless sensor networks. ALGORITHMICA. vol. 4, pp. 412-446 ISSN: 0178-4617.
3. AUSIELLO G.; PASCHOS V. T (2006). Reductions, completeness and the hardness of approximability. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, vol. 172; p. 719-739, ISSN: 0377-2217
4. AUSIELLO G.; BONIFACI V; LAURA L (2008). On-line algorithms for the Asymmetric Traveling Salesman Problem. JOURNAL OF DISCRETE ALGORITHMS, vol. 6 (2); p. 290-298, ISSN: 1570-8667
5. AUSIELLO G.; ALLULLI L; BONIFACI V; LAURA L (2008). On the power of Lookahead in On-line Server Routing Problems. THEORETICAL COMPUTER SCIENCE, ISSN: 0304-3975
6. AUSIELLO G.; BONIFACI V; LAURA L (2008). The on-line prize-collecting traveling salesman problem. INFORMATION PROCESSING LETTERS, ISSN: 0020-0190
7. AUSIELLO G, FRANCIOSA P., ITALIANO G. F. (in stampa). Small stretch (alpha, beta)-spanners in the streaming model. THEORETICAL COMPUTER SCIENCE. ISSN: 0304-3975. accepted for publication.
8. AUSIELLO G, FRANCIOSA P., ITALIANO G.F. (2006). Small stretch spanners on dynamic graphs. JOURNAL OF GRAPH ALGORITHMS AND APPLICATIONS. vol. 10, pp. 365-385 ISSN: 1526-1719.
9. DONATO D, LAURA L, LEONARDI S., MILLOZZI S. (2007). The Web as a graph: How far we are. ACM TRANSACTIONS ON ALGORITHMS. vol. 7(1), pp. 4-34 ISSN: 1549-6325. . ACM Trans. Internet Techn. 7(1): (2007).
10. KOENEMANN J, LEONARDI S., SCHAEFER G, VAN ZWAM S. (2008). A group-strategyproof cost sharing mechanism for the steiner forest game. SIAM JOURNAL ON COMPUTING. vol. 37(5), pp. 1319-1341 ISSN: 0097-5397.
11. LEONARDI S., RAZ D. (2007). Approximating total flow time on parallel machines. JOURNAL OF COMPUTER AND SYSTEMS SCIENCES INTERNATIONAL. vol. 73(6), pp. 875-891 ISSN: 1064-2307. J. Comput. Syst. Sci. 73(6): 875-891 (2007).

12. BARTAL Y, FIAT A, LEONARDI S. (2006). *Lower Bounds for On-line Graph Problems with Application to On-line Circuit and Optical Routing*. SIAM JOURNAL ON COMPUTING. vol. 36(2), pp. 354-393 ISSN: 0097-5397.
13. RAUCH HENZINGER M, LEONARDI S. (2003). *Scheduling multicasts on unit-capacity trees and meshes*. JOURNAL OF COMPUTER AND SYSTEMS SCIENCES INTERNATIONAL. vol. 66(3), pp. 567-611 ISSN: 1064-2307.
14. DEMETRESCU C., ITALIANO G. F. (2006). *Experimental Analysis of Dynamic All Pairs Shortest Path Algorithms*. ACM TRANSACTIONS ON ALGORITHMS. vol. 2, pp. 578-601 ISSN: 1549-6325.
15. DEMETRESCU C., MIKKEL THORUP, REZAUL ALAM CHOWDHURY, VIJAYA RAMACHANDRAN. (2008). *Oracles for distances avoiding a failed node or link*. SIAM JOURNAL ON COMPUTING. vol. 37(5), pp. 1299-1318 ISSN: 0097-5397.
16. DEMETRESCU C., ITALIANO G. F. (2006). *Dynamic shortest paths and transitive closure: Algorithmic techniques and data structures*. JOURNAL OF DISCRETE ALGORITHMS. vol. 4, pp. 353-383 ISSN: 1570-8667.
17. DEMETRESCU C., ITALIANO G. F. (2006). *Fully Dynamic All Pairs Shortest Paths with Real Edge Weights*. JOURNAL OF COMPUTER AND SYSTEM SCIENCES. vol. 72, pp. 813-837 ISSN: 0022-0000.
18. DEMETRESCU C., ITALIANO G. F. (2005). *Trade-Offs for Fully Dynamic Reachability on DAGs: Breaking Through the  $O(n^2)$  Barrier*. JOURNAL OF THE ASSOCIATION FOR COMPUTING MACHINERY. vol. 52(2), pp. 147-156 ISSN: 0004-5411.
19. DEMETRESCU C., ITALIANO G. F. (2004). *A New Approach to Dynamic All Pairs Shortest Paths*. JOURNAL OF THE ASSOCIATION FOR COMPUTING MACHINERY. vol. 51(6), pp. 968-992 ISSN: 0004-5411.
20. BECCHETTI L., CASTILLO C, DONATO D, BAEZA-YATES R, LEONARDI S. (2008). *Link analysis for web s detection*. ACM TRANSACTIONS ON THE WEB. vol. 2, pp. 1-42 ISSN: 1559-1131.
21. BECCHETTI L., KORTEWEG PETER, MARCHETTI-SPACCAMELA, MARTIN SKUTELLA ALBERTO, VITALETTI ANDREA. (2008). *Latency Constrained Aggregation in Sensor Networks*. ACM TRANSACTIONS ON ALGORITHMS. ISSN: 1549-6325. Final version accepted for publication.
22. AUSIELLO G.; BAZGAN C.; DEMANGE M.; PASCHOS V. T. (2005). *Completeness in Differential Approximation Classes*. INTERNATIONAL JOURNAL OF FOUNDATIONS OF COMPUTER SCIENCE, vol. 16; p. 1267-1295, ISSN: 0129-0541

## B) Pubblicazioni di volumi o saggi in volume

1. AUSIELLO G.; V. TH. PASCHOS (2005). *Réductions préservant l'approximabilité*. In: VANGELIS TH. PASCHOS. *Optimisation combinatoire 2. Concepts avancés*. p. 59-90, PARIS: Hermes, ISBN/ISSN: 2-7462-1039-8
2. LUCA ALLULLI, D'AMORE F., ENRICO PUDDU. (2007). *A Platform for Engineering Cache-Oblivious Algorithms and Data Structures: The Architecture*. In: GIUSEPPE F. ITALIANO, PETRA MUTZEL, PETER SANDERS, MARTIN SKUTELLA. *Algorithm Engineering*. Oberwolfach Report No. 25/2007.
3. AUSIELLO G.; V. T. PASCHOS (2006). *Differential ratio approximation*. In: TEOFILLO F. GONZALES. *Approximation Algorithms and Metaheuristics*. Chapman & Hall
4. AUSIELLO G.; L. BECCHETTI (2005). *Algorithmes 'on-line'*. In: VANGELIS TH. PASCHOS. *Optimisation combinatoire 2. Concepts avancés*. p. 249-286, PARIS: Hermes, ISBN/ISSN: 2-7462-1039-8

## C) Pubblicazioni su atti di convegni e congressi

1. AUSIELLO G.; C. BAZGAN; M. DEMANGE; V. PASCHOS (2003). *Completeness in Differential Approximation Classes*. In: *Lecture Notes in Computer Science*, August 2003 Springer Verlag, vol. 2747, p. 179-188
2. E.TREVISANI, VITALETTI A. (2004). *Cell-ID Location Technique, Limits and Benefits: An Experimental Study*. In: *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04)*. WMCSA. (pp. 51-60). ISBN/ISSN: 0-7695-2258-0. WASHINGTON, DC: IEEE Computer Society (UNITED STATES).
3. ALLULLI L; AUSIELLO G.; L. LAURA (2005). *On the Power of Lookahead in On-line Vehicle Routing Problems*. In: *Lecture Notes in Computer Science*. Jeju, August 2005 Springer Verlag, vol. 3595, p. 728-736
4. AUSIELLO G.; B. ESCOFFIER; J. MONNOT; V. TH. PASCHOS (2006). *Reoptimization of minimum and maximum traveling salesman's tours*. In: *Lecture Notes in Computer Science*. Riga, July 2006 Springer Verlag, vol. 4059, p. 196-207, ISBN/ISSN: 3-540-35753-X
5. AUSIELLO G.; BONIFACI V; LAURA L (2005). *On-line algorithms for the Asymmetric Traveling Salesman Problem*. In: *Lecture Notes in Computer Science*. Toronto, August 2005 Springer Verlag, vol. 3608, p. 306-317
6. AUSIELLO G, DEMETRESCU C, FRANCIOSA P., ITALIANO G.F, RIBICHINI A. (2007). *Small Stretch Spanners in the Streaming Model: New Algorithms and Experiments*. In: *Proceedings of ESA 2007. Algorithms - ESA 2007, 15th Annual European Symposium*. October 8-10, 2007. (vol. 4698, pp. 605-617). ISBN/ISSN: 978-3-540-75519-7. : Springer-Verlag.
7. DEBORA DONATO, LEONARDI S., PANAYIOTIS TSAPARAS. (2005). *Stability and Similarity of Link Analysis Ranking Algorithms*. In: *Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005*. ICALP 2005. : Springer-Verlag.
8. JOCHEN KOENEMANN, LEONARDI S., GUIDO SCHAEFER. (2005). *A group-strategyproof mechanism for Steiner forests*. In: *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005*. Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms. (pp. 612-619). VANCOUVER.
9. DEMETRESCU C., FARUOLO P, ITALIANO G. F, THORUP M. (2006). *Does Path Cleaning Help in Dynamic All-Pairs Shortest Paths?*. *Proceedings of the 14th Annual European Symposium on Algorithms (ESA'06)*. Settembre 2006. (pp. 732-743). : Springer Verlag.
10. DEMETRESCU C., FINOCCHI I., RIBICHINI I. (2006). *Trading off space for passes in graph streaming problems*. *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'06)*. (pp. 714-723). : ACM/SIAM (UNITED STATES).
11. DEMETRESCU C., EMILIOZZI S, ITALIANO G.F. (2004). *Experimental Analysis of Dynamic All Pairs Shortest Path Algorithms*. *Proc. of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'04)*. Gennaio 2004. (pp. 369-378).
12. FERRARO PETRILLO U., I. FINOCCHI, G. F. ITALIANO. (2006). *The Price of Resiliency: A Case Study on Sorting with Memory Faults*. In: *Algorithms - ESA 2006, 14th Annual European Symposium, Zurich, Switzerland, September 11-13, 2006*. Algorithms - ESA 2006, 14th Annual European Symposium. 11--13 Settembre 2006. (vol. 4168, pp. 768-779). ISBN/ISSN: 3-540-38875-3. : Springer.
13. L. AULETTA, G. CATTANEO, L. CATUOGNO, A. DE GREGORIO, FERRARO PETRILLO U., A. FRESA, D. PARENTE, G. PERSIANO. (2003). *Applying WBEM to heterogeneous TLC Network Management: an evaluation*. IEEE SoftCOM 2003. 7--10 Ottobre 2003. (pp. 380-385).
14. BECCHETTI L., PETER KORTEWEG, ALBERTO MARCHETTI-SPACCAMELA, MARTIN SKUTELLA, LEEN STOUGIE, ANDREA VITALETTI. (2006). *Latency Constrained Aggregation in Sensor Networks*. In: *Proceedings of the European Symposium on Algorithms (ESA)*. European Symposium on Algorithms (ESA). 2006. : Springer Verlag.
15. VINCENZO BONIFACI, PETER KORTEWEG, MARCHETTI SPACCAMELA A., LEEN STOUGIE. (2008). *Minimizing Flow Time in the Wireless Gathering Problem*. *Symposium on Theoretical Aspects of Computer Science*. Febbraio, 2008. (pp. 109-120). : Internationales Begegnungs- Dagstuhl (GERMANY).
16. L.BURIOL, G. FRAHLING, S. LEONARDI, MARCHETTI SPACCAMELA A., C. SOHLER. (2006). *Counting triangles in data stream*. In: *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. ACM PODS. 26-28 Giugno 2006. (pp. 253-262). ISBN/ISSN: 1-59593-318-2. : ACM Press (UNITED STATES). ACM, Press 2006.
17. M. MASTROGIOVANNI, C. PETRIOLI, M. ROSSI, VITALETTI A., M. ZORZI. (2006). *Integrated data delivery and interest dissemination techniques for wireless sensor networks*. GLOBECOM. November 2006. : IEEE.
18. AUSIELLO G.; V. BONIFACI; L. LAURA (2004). *On explorers, chasers, and cameramen*. In: FUN 2004, p. 287-294

## D) Altro (pubblicazioni non previste nei punti precedenti)

### 4. Richiesta di finanziamento del progetto

#### Note (specificare in dettaglio le spese)

4.1 A) Totale spese per l'acquisto di apparecchiature scientifiche	€ 4.000	2 computer portatili Mac Book Pro o Mac Book Air
4.2 B) Spese generali per la ricerca	€ 0	
4.2.1 Materiali di consumo e manutenzione strumenti (specificare il tipo di materiale e la strumentazione utilizzata)		
4.2.2 Missioni - Seminari	€ 18.000	Missioni a congressi per la presentazione dei risultati della ricerca. Pagamento quote di iscrizione a congressi. Invito di professori visitatori.
4.2.3 Raccolta, codifica e elaborazioni dati	€ 0	
4.2.4 Altre voci: Servizi	€ 12.000	Pagamento di servizi per: - sviluppo software, - gestione sito web, - organizzazione di convegni per la diffusione dei risultati

TOTALE A+B 34.000

4.3 C) Collaborazioni di ricerca (l'importo fisso di 1.550 €, - lorde al mese, per un max di 12 mesi) € 18.600 Collaborazione di ricerca finalizzata alla progettazione, ingegnerizzazione e sperimentazione di algoritmi per grafi di grandissime dimensioni

### 4.4 Ultimi tre anni di finanziamenti ottenuti per ex Progetti di Ateneo

2004: Nessun finanziamento

---

Anno	Fondo assegnato	Fondo non ancora utilizzato
4.4.2 2005	Voce A 2.850	Voce A 161
Â	Voce B 10.450	Voce B 520
Â	Voce C 0	Voce C 0

---

Anno	Fondo assegnato	Fondo non ancora utilizzato
4.4.3 2006	Voce A 3.000	Voce A 3.000
Â	Voce B 9.000	Voce B 5.079
Â	Voce C 9.300	Voce C 0

---

#### **MOTIVAZIONI.**

*Il progetto di ricerca finanziato negli anni 2005 e 2006 aveva per tema 'Algoritmi evoluti per Internet e per il Web'.*

*Per lo svolgimento del programma di ricerca stesso e' stata utilizzata, per una durata di 12 mesi, la collaborazione dell'Ing. Luigi LAURA. La scelta e' stata determinata dalla competenza ed esperienza acquisita dallo stesso durante gli studi ed in particolare durante il dottorato di ricerca, sulle tematiche relative ai modelli e agli algoritmi per il Web.*

#### **CURRICULUM.**

*L'Ing. Luigi LAURA si e' laureato in Ingegneria Informatica presso l'Universita' di Roma La Sapienza nel 1998 con una tesi su Algoritmi on-line per il problema del 'traveling repairman' (relatore il Prof. Ausiello). Nel 2002 l'Ing. LAURA ha conseguito il Dottorato in Ingegneria Informatica. La tesi di Dottorato ha riguardato 'modelli ed algoritmi per il Web'. Dal 2003 ad oggi l'Ing. LAURA ha svolto ricerca con contratti di collaborazione con il Dipartimento di Informatica e Sistemistica ed e' stato docente a contratto di vari corsi presso la Facolta' di Ingegneria. Attualmente e' professore a contratto di Fondamenti di Informatica, presso la stessa Facolta'. Le sue ricerche hanno riguardato prevalentemente il progetto e l'analisi di algoritmi on-line, lo studio di proprieta' e modelli del grafo del Web, lo studio di algoritmi su memoria secondaria per grafi di grandissime dimensioni (come appunto porzioni significative del grafo del Web), lo studio di tecniche di visualizzazione di strutture di dati e programmi.*

#### **PUBBLICAZIONI PRECEDENTI ALLO SVOLGIMENTO DEL PROGETTO**

*L. Laura, S. Leonardi, G. Caldarelli, P. De Los Rios, A multi-layer model of the Web graph. 2nd International Workshop on Web dynamics, Honolulu, Hawaii, May 2002*

*S. O. Krumke, L. Laura, M. Lipmann, A. Marchetti-Spaccamela, W. de Paepe, D. Poensgen, L. Stougie, Non-abusiveness Helps: An  $O(1)$ -Competitive Algorithm for Minimizing the Maximum Flow Time in the Online Traveling Salesman Problem. APPROX 2002, LNCS, Springer, (2002) 200-214*

*L. Laura, S. Leonardi, S. Millozzi, U. Meyer, J. F. Sibeyn, Algorithms and Experiments for the Webgraph. ESA 2003, LNCS, Springer (2003) 703-714*

*D. P. Dubhashi, L. Laura, A. Panconesi, Analysis and Experimental Evaluation of a Simple Algorithm for Collaborative Filtering in Planted Partition Models, FSTTCS 2003, LNCS, Springer (2003), 168-182*

*G. Ausiello, M. Demange, L. Laura, V. Th. Paschos, Algorithms for the On-Line Quota Traveling Salesman Problem. Inf. Process. Lett. 92(2): 89-94 (2004)*

*V. Bonifaci, C. Demetrescu, I. Finocchi, L. Laura, A Java-based system for building animated presentations over the Web. Sci. Comput. Program. 53(1): 37-49 (2004)*

#### **RELAZIONE SULL'ATTIVITA' SVOLTA NELL'AMBITO DEL PROGETTO**

*Nell'ambito del progetto, l'Ing. LAURA ha proseguito lo studio, la progettazione e l'analisi di algoritmi per il Web. In particolare ha curato la realizzazione di algoritmi su memoria secondaria per analizzare struttura e proprieta' di frammenti del grafo del Web (tratti dal repository WebBase) dotati di milioni di nodi e decine di milioni di link. Sempre con riferimento al Web, l'Ing LAURA ha anche studiato aspetti algoritmici della ricerca di informazioni nel Web con l'uso di tecniche algebriche di Information Retrieval. Infine ha studiato problemi di teoria dei giochi derivanti dalla applicazione di tale teoria alla gestione di reti. Tutti i contributi sono stati accettati su riviste di ottimo livello e sono ancora in fase di pubblicazione.*

#### **PUBBLICAZIONI RISULTANTI DALLA COLLABORAZIONE AL PROGETTO**

*Debora Donato, Luigi Laura, Stefano Leonardi, Stefano Millozzi*

*Modeling the Webgraph: how far we are*

*ACM Transaction on Internet Technology (TOIT), to appear.*

*Debora Donato, Luigi Laura, Stefano Leonardi, Ulrich Meyer, Stefano Millozzi, Jop Frederik Sibeyn*

*Algorithms and Experiments for the Webgraph*

*Journal of Graph Algorithms and Applications (World Scientific Publishing), to appear.*

*Vincenzo Bonifaci, Ugo Di Iorio, Luigi Laura*

*The complexity of uniform Nash equilibria and related regular subgraph problems*

*Theoretical Computer Science, 2008, to appear.*

## **4.5 Consuntivo scientifico per l'ultimo anno di finanziamento ottenuto (risultati e pubblicazioni relative)**

*Durante il primo anno di finanziamento i problemi descritti nel progetto sono stati affrontati grazie al lavoro coordinato di tutti i partecipanti. I risultati ottenuti hanno dato luogo ad articoli apparsi su atti di congressi internazionali di primo piano e in gran parte accettati su riviste prestigiose.*

*Sono citate nel seguito le pubblicazioni relative alle direzioni di ricerca proposte.*

#### **MODELLI DI CALCOLO PER GRANDI MOLI DI DATI**

*Sono state definite tecniche di simulazione PRAM nel modello W-Stream [DEM07], sono stati proposti nuovi algoritmi di streaming basati su modelli di streaming tradizionali [ADFIR07], ed e' stata realizzata una piattaforma di sperimentazione di algoritmi cache-oblivious [ADP07]*

#### **METODOLOGIE E STRUMENTI DI VISUALIZZAZIONE DI ALGORITMI PER GRANDI MOLI DI DATI**

*Sono stati realizzati strumenti per la visualizzazione e l'ingegnerizzazione di algoritmi per grandi moli di dati [BDFL06]*

#### **ALGORITMI PER GRAFI DI GRANDI DIMENSIONI**

*Sono stati progettati algoritmi particolarmente efficienti per il calcolo di spanners di grafi in un contesto di data streaming [AFI08], le prestazioni di questi algoritmi sono state confrontate sperimentalmente con algoritmi noti in letteratura [ADIR07, ADFIR08].*

*In [DFR08] sono state studiate tecniche generali per la soluzione di vari problemi su grafi su modelli di data streaming.*

*Nuovi algoritmi cache-oblivious per alcuni problemi fondamentali su grafi sono stati presentati in [ALZ07].*

*Algoritmi in semi-streaming per il calcolo del numero di triangoli incidenti su ciascun vertice in un grafo sono stati proposti in [BBCG08], mentre in [BCDBL08] sono presentati algoritmi basati sulla struttura dei link del Web per la individuazione di Web Spam.*

*Il problema della raccolta di pacchetti in reti wireless e' stato affrontato in [BKMS08a, BKMS08b, BKMS08c], fornendo algoritmi centralizzati e distribuiti.*

#### **RIFERIMENTI BIBLIOGRAFICI DELLE PUBBLICAZIONI RELATIVE ALL'ULTIMO ANNO DI PROGETTO**

- [ADP07] L. Allulli, F. D'Amore, E. Puddu.  
*A Platform for Engineering Cache-Oblivious Algorithms and Data Structures: The Architecture.*  
 In: G. F. Italiano, P. Mutzel, P. Sanders, M. Skutella: Algorithm Engineering, Oberwolfach Report No. 25/2007.
- [ALZ07] L. Allulli, P. Lichodziejewski, N. Zeh.  
*Improved Cache-Oblivious Single-Source Shortest Paths For Small Edge Weights.*  
 in Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07), 2007.
- [ADFIR07] G. Ausiello, C. Demetrescu, P. G. Franciosa, G. F. Italiano, A. Ribichini.  
*Small Stretch Spanners in the Streaming Model: New Algorithms and Experiments*  
 in Proceedings of the 15th Annual European Symposium on Algorithms (Engineering and Applications Track), pp. 605-617, 2007.
- [ADFIR08] G. Ausiello, C. Demetrescu, P. G. Franciosa, G. F. Italiano, A. Ribichini  
*Graph Spanners in the Streaming Model: an Experimental Study.*  
 Algorithmica, 2008. To appear.
- [AFI08] G. Ausiello, P. G. Franciosa, G. F. Italiano.  
*Small stretch (alpha, beta)-spanners in the streaming model*  
 Theoret. Comput. Sci., 2008, in press.
- [BBCG08] L. Becchetti, P. Boldi, C. Castillo, A. Gionis.  
*Efficient Semi-Streaming Algorithms for Local Triangle Counting in Massive Graphs*  
 To appear in the Proceedings of the 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2008.
- [BCDBL08] L. Becchetti, C. Castillo, D. Donato, R. Baeza-Yates, S. Leonardi.  
*Link Analysis for Web Spam Detection*  
 ACM Transactions on the Web, ACM Trans. Web 2 (1), 1-42, 2008.
- [BDFL06] V. Bonifaci, C. Demetrescu, I. Finocchi, L. Laura (2006).  
*Visual Editing of Animated Algorithms: the Leonardo Web Builder.*  
 Proceedings of the 8th International Working Conference on Advanced Visual Interfaces (AVI 2006). : ACM Press.
- [BKMS08a] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, L. Stougie  
*Minimizing flow time in the wireless gathering problem*  
 In Proc. 25th Symposium on Theoretical Aspects of Computer Science, pages 109-120. IBFI Dagstuhl, 2008.
- [BKMS08b] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, L. Stougie.  
*The distributed wireless gathering problem*  
 In Proc. 4th Conf. on Algorithmic Aspects of Information Management, to appear.
- [BKMS08c] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, L. Stougie  
*An approximation algorithm for the wireless gathering problem*  
 Operations Research Letters, to appear.
- [DEM07] C. Demetrescu, B. Escoffier, G. Moruz and A. Ribichini  
*Adapting Parallel Algorithms to the W-Stream Model, with Applications to Graph Problems*  
 In Proceedings of the 32nd International Symposium on Mathematical Foundations of Computer Science (MFCS '07), pages 194-205, 2007.
- [DFR08] C. Demetrescu, I. Finocchi, A. Ribichini  
*Trading off space for passes in graph streaming problems*  
 ACM Transactions on Algorithms, 2008. To appear.

## Finanziamenti P.R.I.N.

	2004	2005	2006	2007
<b>Area</b>	-	09	-	09
<b>Presentati</b>	SI	NO	SI	NO
<b>Finanziati</b>	SI	NO	SI	

Contestualmente alla domanda di Ateneo, il proponente sta presentando anche domanda per Ricerca di Ateneo Federato?  SÌ  NO

## 5. Parere del Dipartimento/Centro di appartenenza del responsabile

Á Data delibera: 19/06/2008 Á Parere: POSITIVO

Firma .....

Data 24/06/2008 16:33