

Massimo Santini

Recapito ufficio:

Dipartimento di Scienze dell'Informazione
Via Comelico, 39/41
20135 Milano
Telefono: 02 503 16259, Fax: 02 503 16373
Email: santini@dsi.unimi.it
Web: <http://santini.dsi.unimi.it/>

Recapito abitazione:

C.so Sempione, 15/a
I-20145 Milano
Telefono: 02 36 511 780, Fax: 02 700 555 449

Curriculum vitae

Sono nato a Milano l'8 settembre 1970.

Nel 1995 ho conseguito la laurea in Scienze dell'Informazione, presso l'Università degli Studi di Milano, con una tesi dal titolo: “*Approximation Algorithms for the Ground State Problem on Tridimensional Ising Spin Glasses*” (relatori: Alberto Bertoni e Paola Campadelli; voto finale: 110/110, *magna cum laude*).

Nel 2000 ho conseguito il dottorato di ricerca in Informatica, presso l'Università degli Studi di Milano, con una tesi dal titolo: “*Random Generation and Approximate Counting of Combinatorial Structures*” (relatori: Alberto Bertoni, Bruno Codenotti e Massimiliano Goldwurm).

Negli anni 1999–2000 sono stato titolare di un assegno di ricerca dal titolo “Tecniche genetiche nella previsione per applicazioni” presso il Dipartimento di Scienze dell'Informazione dell'Università di Crema (responsabile: Gianni Degli Antoni).

Nel 2001 (nel semestre da gennaio a giugno) sono stato collaboratore alla ricerca presso l'Istituto per le Applicazioni della Telematica del Consiglio Nazionale delle Ricerche. Quindi (nei mesi di luglio e agosto) sono stato assunto come *research associate*, alla Thayer School of Engineering, presso il Dartmouth College (USA). Negli anni 2003–2006 il rapporto con il Consiglio Nazionale delle Ricerche si è consolidato in un incarico di ricerca a titolo gratuito dell'Istituto di Informatica e Telematica.

Negli anni 2001–2004 sono stato ricercatore, nel settore scientifico-disciplinare INF/01–Informatica, presso il Dipartimento di Scienze Sociali Cognitive e Quantitative dell'Università di Modena e Reggio Emilia. Nel semestre da marzo ad agosto del 2003 sono stato inoltre ospite come *visiting scholar* presso il Computer Science Department della Brown University (USA).

Dal 2004 sono ricercatore confermato, nel settore scientifico-disciplinare INF/01–Informatica, presso il Dipartimento di Scienze dell'Informazione dell'Università degli Studi di Milano, dove afferisco al *Laboratorio di Algoritmica per il Web (LAW)*.

Attività di ricerca

I miei interessi di ricerca riguardano la progettazione e l'analisi degli algoritmi.

Complessità strutturale. Durante la tesi di laurea (e nel primo periodo del dottorato) mi sono occupato del problema di approssimazione dell'energia dello stato fondamentale per vetri di spin nel modello di Ising [J1, C1, T3, M2]. Tale problema è formulabile come problema di ottimizzazione quadratica su $\{0, 1\}$ ed è notoriamente classificato come NP-difficile; i principali risultati conseguiti riguardano il caso di strutture in cui $n > 0$ atomi

giacciono su due livelli. Per tale caso è stato dato un algoritmo polinomiale di approssimazione con errore $O(n/\lg n)$ ed è stato provato un limite inferiore $\Omega(n/\lg^2 n)$ all'errore di approssimazione per qualunque algoritmo polinomiale (modulo una naturale congettura di complessità strutturale).

Conteggio e generazione casuale di strutture combinatorie. Il dottorato di ricerca ha poi riguardato principalmente l'ambito dei linguaggi formali; in particolare, mi sono occupato di algoritmi per la generazione casuale uniforme delle parole di linguaggi *context-free* con grado di ambiguità limitato [J3, C3, C4, T5, M3]. Nel caso dei linguaggi finitamente ambigui, il principale risultato conseguito è un algoritmo che genera parole di lunghezza n in tempo $O(n^2 \log n)$ su RAM con costo logaritmico; per il caso di linguaggi polinomialmente ambigui, si è mostrato che il problema è risolvibile con circuiti booleani di una opportuna versione randomizzata di NC^2 .

Mi sono inoltre occupato di una questione relativa a problemi di conteggio in linguaggi traccia [J2, T6]. In tale contesto, il risultato ottenuto è la dimostrazione dell'unicità della radice di modulo minimo dei polinomi di *clique*, fatto che consente di ottenere una valutazione asintotica di alcune sequenze legate sia a problemi enumerativi che allo studio di algoritmi riguardo ai monoidi traccia.

Il risultato più recente in questo ambito riguarda i permutomini, una particolare classe di poliomini definita a partire da una permutazione. Di tale struttura combinatoria è stato risolto il problema del conteggio [J11, C15, T12], dando la prima formula chiusa in letteratura per il numero di permutomini di dimensioni assegnate.

Da un punto di vista più applicativo [T14], si è iniziata una analisi dei linguaggi di alberi, con particolare riferimento alle questioni dell'ambiguità e del determinismo, di particolare rilevanza nel contesto dell'XML.

Modelli ed algoritmi per il web. Dopo il periodo del dottorato, i miei interessi di ricerca si sono spostati su alcuni aspetti algoritmici legati al *world wide web*. Una prima intensa attività di ricerca [J5, C6, C7, C9] ha riguardato la progettazione e la realizzazione di un *crawler* distribuito e *fault tolerant*. Questa attività è coincisa con la nascita di due laboratori di algoritmica per il web, uno presso il Dipartimento di Scienze dell'Informazione di Milano, e l'altro presso l'Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche, a Pisa.

Il *software* realizzato durante questa fase, documentato nell'attività pubblicistica citata, ha messo i due laboratori in grado di raccogliere ed elaborare una grande mole di dati, nell'ordine di centinaia di milioni di documenti web. Inoltre ha favorito da un lato lo svilupparsi di una collaborazione internazionale con l'Università di Nagaoka (Giappone) –nell'ambito di un progetto finanziato dall'UNESCO per lo studio della diffusione delle lingue minori sul web [C13]– e ha determinato dall'altro la partecipazione del LAW al progetto europeo DELIS.

Nell'ambito di tale progetto, recentemente conclusosi, l'uso di UbiCrawler è stato fondamentale nella raccolta di dodici "istantanee" mensili di larga scala del web inglese che costituiscono il primo *corpus* pubblico di pagine web *annotato temporalmente* [J10]; parte di tale collezione è stata utilizzata nel costituire il primo repertorio pubblico [J9, T11] di dati per la ricerca nell'ambito dell'*adversarial information retrieval*, con particolare riferimento al *web spam*.

I dati ricavati tramite il *crawler* hanno consentito sia analisi sperimentali di porzioni rilevanti del web [C8], che suggerito nuovi problemi algoritmici [J6, T10]. Il principale risultato a questo riguardo è dato da un algoritmo approssimato per l'efficiente memorizzazione ed interrogazione di *corpus* documentali organizzati gerarchicamente, tipici dei repertori di pagine web.

Un ulteriore filone di ricerca in questo ambito è legato allo studio del PageRank, un indice di qualità dei documenti web. In tale contesto, il primo risultato ottenuto ha riguardato un'analisi empirica dell'influenza che l'ordine in cui le pagine web vengono raccolte ha sul calcolo di tale indice [J7, C11]; una ulteriore analisi empirica ha riguardato invece il comportamento di altri indici di qualità legati a PageRank, mostrando in particolare [C14] che il calcolo pratico di tali indici può celare problemi non indifferenti di stabilità e precisione numerica.

Da un punto di vista più teorico [C12], si è invece studiato analiticamente il comportamento del PageRank in funzione del *damping factor* (uno dei parametri fondamentali in base al quale viene definito), ottenendo una

espressione in forma chiusa per le derivate di PageRank (rispetto a tale parametro e di qualsiasi ordine) ed un algoritmo efficiente per la loro approssimazione numerica.

Una ulteriore analisi [J8] ha consentito di mettere in luce una relazione tra una particolare classe di catene di Markov, di cui il PageRank è un caso particolare, con le fibrazioni di grafi. Tale risultato mostra come, sotto opportune condizioni, il PageRank sia costante all'interno della stessa fibra, un fatto che può essere utilizzato, tra l'altro, per risolvere in modo efficiente il problema dell'isomorfismo in una particolare classe di grafi.

Miscellanea. Infine, alcuni interessi secondari di ricerca hanno riguardato (in ordine sparso): una questione di algoritmica ricreativa ispirata ad un problema di misurazione descritto da Nicolò Tartaglia [J4, C2, T2], lo studio del miglioramento adattativo di algoritmi per l'apprendimento computazionale [T1, T4, M1], l'analisi "non standard" dal punto di vista della teoria dei modelli [T7] e un'applicazione dell'XML alla rappresentazione dei lessici tecnici multilingue [T13]. Durante il godimento dell'assegno di ricerca mi sono occupato di algoritmi genetici per la previsione di serie finanziarie [C5, T8] e, nel periodo in cui ho lavorato per il Dartmouth College, mi sono occupato di controllo distribuito [C10, T9].

Publicazioni scientifiche

- [J1] Roberto Posenato and Massimo Santini. A new lower bound on approximability of the ground state problem for tridimensional ising spin glasses. *Information Processing Letters*, 68(4):167–171, 1998. GRIN=B,IF=0.196.
- [J2] Massimiliano Goldwurm and Massimo Santini. Clique polynomials have a unique root of smallest modulus. *Information Processing Letters*, 75(3):127–132, 2000. GRIN=B,IF=0.360.
- [J3] Alberto Bertoni, Massimiliano Goldwurm, and Massimo Santini. Random generation for finitely ambiguous context-free languages. *Revue d'Automatique, d'Informatique et de Recherche Opérationnelle–Theoretical Informatics Applications*, 35(6):499–512, 2001. GRIN=B,IF=0.154.
- [J4] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Measuring with jugs. or: What if mathematicians were asked to defuse bombs. *Theoretical Computer Science*, 282(2):259–270, 2002. GRIN=A,IF=0.417.
- [J5] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. UbiCrawler: a scalable fully distributed web crawler. *Software Practice and Experience*, 34(8):711–726, 2003. GRIN=A,IF=0.718.
- [J6] Bruno Codenotti, Gianluca De Marco, Mauro Leoncini, Manuela Montangero, and Massimo Santini. Approximation algorithms for a hierarchically structured bin packing problem. *Information Processing Letters*, 75(3):215–221, 2004. GRIN=B,IF=0.453.
- [J7] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Paradoxical effects in PageRank incremental computations. *Internet Mathematics*, 2(3):387–404, 2005.
- [J8] Paolo Boldi, Violetta Lonati, Massimo Santini, and Sebastiano Vigna. Graph fibrations, graph isomorphism, and PageRank. *Revue d'Automatique, d'Informatique et de Recherche Opérationnelle–Theoretical Informatics Applications*, 40(3):227–253, 2006. GRIN=B,IF=0.692.
- [J9] Carlos Castillo, Debora Donato, Luca Becchetti, Paolo Boldi, Massimo Santini, and Sebastiano Vigna. A reference collection for web spam detection. *ACM SIGIR Forum*, 40(2):11–24, 2006.
- [J10] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. A large time-aware graph. *ACM SIGIR Forum*, 42(1):78–83, 2008.
- [J11] Paolo Boldi, Violetta Lonati, Roberto Radicioni, and Massimo Santini. The number of convex permutominoes. *Information and Computation*, 206(9–10):1074–1083, 2008. GRIN=A,IF=0.983.

- [J12] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. PageRank: Functional dependencies. *ACM Transactions on Information Systems*, 27(4):1–23, 2009. GRIN=A,IF=1.969.
- [J13] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Permuting web and social graphs. *Internet Mathematics*, 6(3):257–283, 2009.
- [J14] Paolo Boldi, Andrea Marino, Massimo Santini, and Sebastiano Vigna. Bubing: Massive crawling for the masses. *ACM Transactions on The Web*, 12:12:1–12:26, 2018.
- [C1] Alberto Bertoni, Paola Campadelli, Roberto Posenato, and Massimo Santini. Approximability of ground state problem on tridimensional ising spin glasses. In Alfredo De Santis, editor, *5th Italian Conference on Theoretical Computer Science*, pages 492–496, Ravello, Italy, 1995. World Scientific.
- [C2] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Measuring with jugs. Or: What if mathematicians were asked to defuse bombs. In Elena Lodi, Linda Pagli, and Nicola Santoro, editors, *Fun with Algorithms. Proceedings of the International Conference*, number 4 in Proceedings in Informatics, pages 52–64, Elba, Italy, 1999. Carleton Scientific.
- [C3] Alberto Bertoni, Massimiliano Goldwurm, and Massimo Santini. Random generation and approximate counting of ambiguously described combinatorial structures. In Horst Reichel and Sophie Tison, editors, *Proceedings of 17th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, number 1770 in Lecture Notes in Computer Science, pages 567–580. Springer–Verlag, 2000. GRIN-A.
- [C4] Massimiliano Goldwurm, Beatrice Palano, and Massimo Santini. On the circuit complexity of random generation problems for regular and context-free languages. In Ferreira Afonso and Horst Reichel, editors, *Proceedings of 18th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, number 2010 in Lecture Notes in Computer Science, pages 305–316. Springer–Verlag, 2001. GRIN-A.
- [C5] Massimo Santini and Andrea Tettamanzi. Genetic programming for financial time series prediction. In J. Miller, M. Tomassini, P.L. Lanzi, C. Ryan, A.G.B. Tetamanzi, and W.B. Langdon, editors, *Proceedings of 4th European Conference on Genetic Programming (EuroGP)*, number 2038 in Lecture Notes in Computer Science, pages 361–370. Springer–Verlag, 2001.
- [C6] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. Trovatore: Towards a highly scalable distributed web crawler. In *Poster Proceedings of the 10th International World Wide Web Conference*, Hong Kong, China, 2001. GRIN-B.
- [C7] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. UbiCrawler: Scalability and fault-tolerance issues. In *Poster Proceedings of the 11th International World Wide Web Conference*, Hawaii, USA, 2002. GRIN-B.
- [C8] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. Structural properties of the african web. In *Poster Proceedings of the 11th International World Wide Web Conference*, Hawaii, USA, 2002. GRIN-B.
- [C9] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. UbiCrawler: A scalable fully distributed web crawler. In *Proceedings of the 8th Australian World Wide Web Conference*, Twin Waters Resort, Australia, 2002.
- [C10] Valentino Crespi, George Cybenko, Daniela Rus, and Massimo Santini. Decentralized control for coordinated flow of multi-agent systems. In *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN)*, volume 3 of *Neural Networks, 2002*, pages 2604–2609, Honolulu, Hawaii, 2002. IEEE Press. GRIN-B.
- [C11] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Do your worst to make the best: Paradoxical effects in PageRank incremental computations. In Stefano Leonardi, editor, *Algorithms and Models for the Web-Graph: 3rd International Workshop (WAW2004)*, number 3248 in Lecture Notes in Computer Science, pages 168–180, Rome, Italy, 2004. Springer–Verlag.

- [C12] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. PageRank as a function of the damping factor. In *Proceedings of the 14th International World Wide Web Conference*, pages 557–566, Chiba, Japan, 2005. ACM Press.
- [C13] Yoshiki Mikami, Pavol Zavorsky, Mohd Zaidi Abd Rozan, Irwan Nizan Ayob, Paolo Boldi, Massimo Santini, and Sebastiano Vigna. The language observatory project. In *Poster Proceedings of the 14th International World Wide Web Conference*, Chiba, Japan, 2005. ACM Press. GRIN-B.
- [C14] Paolo Boldi, Roberto Posenato, Massimo Santini, and Sebastiano Vigna. Traps and pitfalls of topic-biased pagerank. In William Aiello, Andrei Broder, Jeannette Janssen, and Evangelos Milios, editors, *Algorithms and Models for the Web-Graph: 4th International Workshop (WAW2006)*, number 4936 in Lecture Notes in Computer Science, pages 107–116, Banff, Canada, 2008. Springer–Verlag.
- [C15] Paolo Boldi, Violetta Lonati, Roberto Radicioni, and Massimo Santini. The number of convex permutominoes. In *Proceedings of the 1st International Conference on Language and Automata Theory and Applications*, volume 14, pages 151–162, Tarragona, Spain, 2007.
- [C16] Ilaria Bordino, Paolo Boldi, Debora Donato, Massimo Santini, and Sebastiano Vigna. Temporal evolution of the uk web. In *Proceedings of the 2008 IEEE International Conference on Data Mining Workshops*. IEEE Computer Society, 2008.
- [C17] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Permuting web graphs. In *Algorithms and Models for the Web-Graph: 6th International Workshop (WAW2009)*, number 5427 in Lecture Notes in Computer Science, pages 116–126, Barcelona, Spain, 2009. Springer–Verlag.
- [C18] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Permuting web and social graphs. In *Proceedings of the 20th International Conference on World Wide Web*, Hyderabad, India, 2011. ACM Press.
- [C19] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Layered label propagation: a multiresolution coordinate-free ordering for compressing social networks. In *Proceedings of the 20th International Conference on World Wide Web*, pages 587–596, New York, 2011. ACM Press.
- [C20] Paolo Boldi, Andrea Marino, Massimo Santini, and Sebastiano Vigna. Bubing: Massive crawling for the masses. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 227–228. ACM, 2014.
- [C21] Paolo Boldi, Corrado Monti, Massimo Santini, and Sebastiano Vigna. Liquid fm: recommending music through viscous democracy. In *Italian Information Retrieval Workshop*, volume 1404, pages 1–12. CEUR, 2015.
- [T1] Massimo Santini. Boosting revisited. Technical Report RI-DSI 205-97, Dipartimento di Scienze dell’Informazione, Milano, 1997.
- [T2] Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Measuring with jugs. or: What if mathematicians were asked to defuse bombs. Technical Report RI-DSI 212-98, Dipartimento di Scienze dell’Informazione, Milano, 1998.
- [T3] Alberto Bertoni, Paola Campadelli, Roberto Posenato, and Massimo Santini. Approximability of ground state problem for tridimensional ising spin glasses. Technical Report RI-DSI 217-98, Dipartimento di Scienze dell’Informazione, Milano, 1998.
- [T4] Massimo Santini. Boosting regression via classification. Technical Report RI-DSI 221-98, Dipartimento di Scienze dell’Informazione, Milano, 1998.
- [T5] Alberto Bertoni, Massimiliano Goldwurm, and Massimo Santini. Random generation and approximate counting of ambiguously described combinatorial structures. Technical Report RI-DSI 236-99, Dipartimento di Scienze dell’Informazione, Milano, 1999.

- [T6] Massimiliano Goldwurm and Massimo Santini. Clique polynomials have a unique root of smallest modulus. Technical Report RI-DSI 247-00, Dipartimento di Scienze dell'Informazione, Milano, 2000.
- [T7] Massimo Santini. Una nota sui modelli nonstandard. Technical Report RI-DSI 248-00, Dipartimento di Scienze dell'Informazione, Milano, 2000.
- [T8] Massimo Santini and Andrea Tettamanzi. Genetic programming for financial time series prediction. Technical Report Nota del Polo 35, Polo Didattico e di Ricerca di Crema, Crema (CR), 2001.
- [T9] Valentino Crespi, George Cybenko, Daniela Rus, and Massimo Santini. Decentralized control for coordinated flow of multi-agent systems. Technical Report TR2002-414, Dartmouth College, Computer Science, Hanover, NH, January 2002.
- [T10] Bruno Codenotti, Gianluca De Marco, Mauro Leoncini, Manuela Montangero, and Massimo Santini. Efficient strategies for partitioning and querying a hierarchical document space. Technical Report IIT TR-25/2002, Istituto di Informatica e Telematica (CNR), Pisa, December 2002.
- [T11] Carlos Castillo, Debora Donato, Luca Becchetti, Paolo Boldi, Stefano Leonardi, Massimo Santini, and Sebastiano Vigna. A reference collection for web spam. Technical Report DELIS-TR-405, Dynamically Evolving, Large-scale Information Systems EU Project, 2006.
- [T12] Paolo Boldi, Violetta Lonati, Massimo Santini, and Roberto Radicioni. The number of convex permutominoes. Technical Report RI-DSI 311-06, Dipartimento di Scienze dell'Informazione, Milano, 2006.
- [T13] Alice Lonati, Violetta Lonati, and Massimo Santini. Modeling and transforming a multilingual technical lexicon for conservation-restoration using XML. Technical Report RI-DSI 317-07, Dipartimento di Scienze dell'Informazione, Milano, 2007.
- [T14] Paolo Boldi, Violetta Lonati, Roberto Radicioni, Massimo Santini, and Sebastiano Vigna. Tree language determinism, ambiguity and typing: towards a uniform approach. Technical Report RI 320-08, Dipartimento di Scienze dell'Informazione, Milano, 2008.
- [T15] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Compressing social networks by community detection. Technical Report RI 330-10, Dipartimento di Scienze dell'Informazione, Milano, 2010.
- [M1] Massimo Santini. *Boosting Revisited*. Minor PhD dissertation, Dipartimento di Scienze dell'Informazione, 1997.
- [M2] Massimo Santini. Algoritmi di approssimazione per il problema dello stato fondamentale dei vetri di spin a tre dimensioni. Master's thesis, Dipartimento di Scienze dell'Informazione, 1995.
- [M3] Massimo Santini. *Random Uniform Generation and Approximate Counting of Combinatorial Structures*. PhD thesis, Dipartimento di Scienze dell'Informazione, 1999.

Partecipazione a progetti

Sono stato membro dei seguenti progetti di ricerca:

- Neural and Computational Learning 2 (NeuroCOLT2), ESPRIT Working Group, 1998;
- Modelli di calcolo innovativi: metodi sintattici e combinatori, COFIN MIUR, 1998;
- Linguaggi formali e automi: teoria e applicazioni, COFIN MIUR, 2001;

- ALINWEB: Algoritmica per Internet e per il Web, COFIN MIUR, 2002;
- DELIS: Dynamically Evolving Large-scale Information Systems, Integrated European Project, “Complex Systems” Proactive Initiative, Sixth Framework Programm (come **responsabile** dell’unità locale);
- Grafi del web e ranking, COFIN MIUR, 2006;
- CRESCO: Centro computazionale di RicErca sui Sistemi COmplessi, Programma Operativo Nazionale “Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione” (come **responsabile** dell’unità locale).
- NADINE: New tools and algorithms for directed network analysis, EC FET Open project, Seventh Framework Programm.

Attività seminariale e presentazioni

Ho tenuto le seguenti presentazioni:

- *Approximability of Ground State Problem on Tridimensional Ising Spin Glasses*, alla “Fifth Italian Conference on Theoretical Computer Science” (Ravello, 1995).
- *Random Generation and Approximate Counting of Ambiguously Described Combinatorial Structures*, al “17th Annual Symposium on Theoretical Aspects of Computer Science” (Lille, Francia, 2000).
- *On the Sequential and Circuit Complexity of the Random Generation Problem for Polynomially Ambiguous Context-free Language*, due seminari al “Laboratoire d’Informatique Algorithmique: Fondements et Applications”, Université Paris VII e al “Laboratoire de Recherche en Informatique”, Université Paris-Sud (Parigi, Francia, 2000).
- *On the Circuit Complexity of Random Generation Problems for Regular and Context-Free Languages*, al “18th Annual Symposium on Theoretical Aspects of Computer Science” (Dresden, Germania, 2001).
- *UbiCrawler: A scalable fully distributed web crawler*, alla “8th Australian World Wide Web Conference” (Twin Waters, Australia, 2002).
- *Web Algorithmics: the Case of Crawling*, ospite della “Algorithms and Cryptography, Seminar Series”, Macquarie University, (Sydney, Australia, 2002).
- *UbiCrawler: A scalable fully distributed web crawler*, ospite del “SIA: Seminario Interdipartimentale di Algoritmica”, Università La Sapienza di Roma (2004).
- *Collecting web data with UbiCrawler*, al meeting annuale COFIN ALINWEB, Università di Pisa (2004).
- *PageRank as a Function of the Damping Factor*, al “1st Bertinoro Workshop on Data-Mining (DaMNn)” (Bertinoro, 2005).
- *The First Step Towards World Wide Web Data Mining*, al meeting annuale DELIS, Max-Planck-Institut Informatik (Saarbrücken, Germany, 2005).
- *WebSnap: Data and Tools*, al meeting annuale DELIS, Università La Sapienza di Roma (2006).

Attività didattica

Nell'ambito delle lauree quinquennali, triennali e specialistiche, ho impartito i seguenti insegnamenti:

- Laboratorio di algoritmi e strutture dati (presso l'Università degli Studi di Milano, negli a.a. 1999/00 e 2000/01),
- Laboratorio di programmazione (presso l'Università dell'Insubria, nell'a.a. 2000/01),
- Metodi per la gestione dei dati (presso l'Università di Modena e Reggio Emilia, negli a.a. 2001/02, 2002/03 e 2003/04),
- Gestione e programmazione per reti e siti (presso l'Università di Modena e Reggio Emilia, negli a.a. 2001/02, 2002/03 e 2003/04),
- Informatica e Laboratorio di informatica (presso l'Università degli Studi di Milano, nell'a.a. 2004/05).
- Programmazione e laboratorio di programmazione (presso l'Università degli Studi di Milano, negli a.a. 2005/06 e 2006/07).
- Laboratorio di programmazione (presso l'Università degli Studi di Milano, dall'a.a. 2007/08 all'a.a. 2017/18).
- Laboratorio di programmazione 1 (presso l'Università degli Studi di Milano, nell'a.a. 2018/19).
- Linguaggi e traduttori (presso l'Università degli Studi di Milano, nell'a.a. 2018/19).

Ho inoltre impartito moduli, o parti monografiche, nell'ambito dei seguenti insegnamenti:

- Algoritmi, algoritmi per reti (presso l'Università *La Sapienza* di Roma, nell'a.a. 2003/04);
- Programmazione (presso il "Master in Elettronica e Sistemi Digitali" dell'Università di Milano, nell'a.a. 2005/06).
- Sviluppo software in gruppi di lavoro complessi (presso l'Università degli Studi di Milano, negli a.a. dal 2016/2017 al 2018/2019).

Negli anni 2011 e 2013, ho tenuto due edizioni del corso di dottorato "Systematic Debugging Techniques" previsto dal curriculum "Advanced Development Techniques" della scuola di dottorato dell'Università degli Studi di Milano. Ho inoltre svolto cicli di lezioni nell'ambito dei seguenti corsi di dottorato:

- WEB BAR: Bertinoro International PhD School (presso il Bertinoro international Center for informatics, nel 2004).
- Algoritmica per il web (presso l'Università degli Studi di Milano, nell'a.a. 2006/07).
- Modelli Markoviani (presso l'Università degli Studi di Milano, nell'a.a. 2006/07).

Ho tenuto corsi di programmazione nell'ambito del primo ciclo del "Tirocinio Formativo Attivo" (TFA) e del primo ciclo dei "Percorsi Abilitanti Speciali" (PAS) della classe A/042, presso l'Università degli Studi di Milano.

Dal 2000 al 2010 sono stato membro del Gruppo Selezionatori Nazionali per le Olimpiadi Internazionali di Informatica (IOI), sotto la supervisione dell'AICA e del Ministero della Pubblica Istruzione.

Attività organizzativa e di servizio

Ho svolto attività come *referee* per le riviste: *Random Structure and Algorithms* e *World Wide Web Journal* e per i convegni: *Symposium on Discrete Algorithms (SODA)*, *Symposium on Theoretical Aspects of Computer Science (STACS)*, *World Wide Web Conference (WWW)*, *ACM International Conference on Web Search and Data Mining (WSDM)*, *ACM Symposium on Applied Computing (SAC)*, *Computational Learning Theory (COLT)* e *Neural Information Processing Systems (NIPS)*.

Sono stato membro del comitato di programma delle conferenze: *World Wide Web Conference (WWW)* e *ACM International Conference on Web Search and Data Mining (WSDM)*. Nel 2019 sono stato *Proceedings Chair* dell'*ACM Conference on Web Science (WebSci)*.

Sono stato membro del collegio docenti del “Dottorato di ricerca in Scienze Sociali Teoriche e Applicate” (con sede presso l’Università di Modena e Reggio Emilia) e sono stato membro del collegio docenti della “Scuola di Dottorato in Informatica” (con sede presso l’Università degli Studi di Milano).

Nell’ambito del mio servizio presso l’Università di Modena e Reggio Emilia, sono stato membro della commissioni “Supplenze e Contratti” e “Polo Informatico e Servizi di Biblioteca” della Facoltà di Scienze della Comunicazione e dell’Economia. Più recentemente, durante il mio servizio presso l’Università di Milano, sono stato membro delle commissioni “Servizi Informatici” e “Web”.

Milano, 31 gennaio 2019

In fede, Massimo Santini