CURRICULUM VITAE AND RESEARCH PLAN
Luca Becchetti

June 10, 2013

Contents

Personal information

- Name: Luca Becchetti
- Birth: Rome, Italy, January 11th, 1966
- Nationality: Italian
- Address: Via Calatafimi 31, Rome, Italy (ZIP code: 00185)
- Position: permanent researcher at Dipartimento di Informatica e Sistemistica, Sapienza University of Rome
- Office: via Ariosto 25, 00185, Roma
- Tel.: +39 06 77274 025
- Fax: +39 06 77274 002
- e-mail: Luca.Becchetti@dis.uniroma1.it

1 Education

- 1995 - 1998: PhD in Computer Science, Sapienza University of Rome (Advisor: Prof. Alberto Marchetti - Spaccamela).
- July 1999 - December 1999: Post Doc at University of Graz, in the combinatorial optimization group, under Prof. Gerhard Woeginger.

2 Current position

March 2001 - present: permanent position as researcher at Dipartimento di Informatica e Sistemistica, Sapienza University of Rome.

3 Languages

Italian (mother tongue), English (fluent, written and spoken), German (fluent, written and spoken).

4 Research plan

My main research directions for the next 3 years are in the following areas:

Decentralized mining and information retrieval tasks in complex networks. Large scale Internet and Web applications continuously perform data mining and information retrieval tasks that benefit both end users and service providers. Performing these tasks poses algorithmic and computational challenges that are the focus of intensive research. Main issues are typically the very large or huge size of the data and the fact that in many cases (e.g., P2P networks) information is spread among the nodes of a possibly very large network of uncoordinated agents, each possessing limited computational resources and an only partial view of the network.

A first line of research will be continuing the study of efficient and scalable algorithmic techniques initiated in [42, 43, 41, 2, 24, 17], with emphasis on solutions that address the issues of decentralization and lack of coordination outlined above, and privacy. This line of research has important ramifications in the area of pervasive and autonomic systems, as recognized by the European Community in the recent past and in the 7th framework programme (e.g., topic 8.5: Self-Awareness in Autonomic Systems, in FET Proactive ICT Call 5). First results in this direction have been achieved in [44, 45, 18, 51, 23, 5, 3, 57].

Optimization in large and complex networks. I plan to continue the study of optimization problems arising in complex systems and social networks, with emphasis on resource allocation and task scheduling. Many information retrieval tasks in large-scale Web applications can be recast as optimization problems over possibly dynamic data sets. An example is the link optimization framework considered in [19] and briefly described in Subsection 5.1. Further challenges arise in Web and Internet applications, where important performance optimization issues, e.g. in Web

search applications, can be recast as on-line combinatorial/stochastic optimization problems. Finally, a novel line of research, initiated in cooperation with Yahoo! Research labs, concerns interesting combinatorial optimization problems arising in the automatic formation of teams to perform tasks arriving off-line or on-line.

5 Summary of current and past research activity

My main current research interests are in the Information Retrieval and Mining of complex networks and massive data sets (Subsection 5.1). Further interests include resource scheduling (Subsection 5.2) and network design and management (Subsection 5.3).

5.1 Web, social networks and massive data sets

This subsection describes past and ongoing research in the mining and analysis of complex networks and massive data sets. This line of research partly arose within ongoing collaboration with Yahoo! research labs, Barcelona.

Web spamming. Given its huge size, interaction with the Web is nowadays mostly mediated by search engines, whose main purpose is assisting users in the retrieval of information that is more likely to meer their expectations. On the ther hand, the Web contains numerous profit-seeking ventures that are attracted by the prospect of reaching millions of users at a very low cost. Therefore, there is an economic incentive in manipulating search engines' listings by creating pages that score high independently of their real merits. In practice, such manipulation is widespread and, in many cases, successful. All deceptive actions that try to increase the ranking of a page in search engines are generally referred to as *Web spam* or *spamdexing*. In the era of textual search engines, Web spamming relied essentially on the manipulation of Web pages' contents. Nowadays, these techniques are complemented by others whose aim is inducing local perturbations of the link structure of the Web, so as to raise the ranking achieved by one or more target pages. These forms of Web spamming (often referred to as *link spamming*) can be very hard to detect.

[42, 43, 41, 2] propose innovative and efficient techniques for the automatic detection of link spamming activity in the Web. Their implementation requires the estimation of statistical indices that reflect the topological structure of the neighbourhood of every node of the Web sample under consideration. Given the large or huge size of the Web samples of interest in real applications, important scalability issues arise. The papers listed above propose new techniques that make this computation feasible and scalable over massive data sets. The solutions proposed have allowed the implementation of automatic classifiers for link Web spamming, whose effectiveness has been experimentally tested on a collection of Web pages described in [55], which is now adopted as a reference benchmark in the community.

Local aggregation properties in social networks. The clustering coefficient measures the extent to which nodes in a network that are adjacent to the same node are also adjacent to each other (roughly speaking, the extent to which my friends are also friends to each other). This index is important in many applications, such as detecting community structure in complex biological systems, social networks or the Web, for example. Given the large or huge size that is

typical of many such systems, its exact computation for every node of the whole network is often unfeasible. [24, 17] propose efficient probabilistic techniques to accurately estimate the local clustering coefficient that scale to networks of large or huge size. In [24, 17] these techniques are applied to the detection of spamming activity in the Web and to the analysis of content quality in virtual communities, specifically the Yahoo! Answers one.

Query recommendation. [19] studies the problem of recommending query reformulations, i.e., suggestions for related queries sometimes appearing at the bottom of search engine result pages. This paper casts the problem in an optimization framework, in which user behaviour is modelled as a random walk over a suitably defined query-flow graph and the goal is to add new links (i.e., query recommendations), so as to optimize an expected utility, where expectation is taken with respect to the perturbed random walk. A strategy is proposed, with provably almost optimal performance in most cases of practical interest. The techniques proposed can be applied to other scenarios where user behavior can be modeled as a Markov chain.

Caching techniques for query answering. The paper [21] introduces the problem of *query covering* as a means to efficiently cache query results. The general idea is to populate the cache with documents that contribute to the result pages of a large number of queries, as opposed to caching the "top" documents for each query. It turns out that the problem is hard and solving it requires knowledge of the "structure" of the queries and the results space, as well as knowledge of the input query distribution. In this paper, the problem is formulated under the framework of online stochastic optimization; theoretically it can be seen as an online stochastic version of set multicover. While the problem is NP-hard to be solved exactly, the paper shows that for any distribution it can be approximated using a simple greedy approach. These theoretical findings are complemented by experimental activity on real datasets, showing the feasibility and potential interest of query-covering approaches in practice.

Streaming algorithms. Maintaining statistics of interest over massive data streams has received considerable attention in the past few years, motivated by a myriad of applications in several areas. [49] considers the problem of maintaining, at every point in time, a possibly accurate estimate of the maximum value over a fixed window W of recent observations when only $k << W$ words of memory are available. Unlike previous approaches, this paper uses competitive analysis to compares the performance of the online streaming algorithm against an optimal adversary that knows the entire sequence in advance. Almost optimal strategies for the problem are proposed and analyzed.

The paper [52] proposes fully decentralized algorithms that answer *locally* continuous aggregate queries on the number of distinct events, total number of events and the second frequency moment in the scenario outlined above. The proposed algorithms use in the worst case or on realistic distributions sublinear space at every node. This paper also proposes strategies that minimize the communication needed to update the aggregates when new events are observed. Finally, experimental analysis is presented, that provides evidence for the efficiency and accuracy of the proposed strategies on realistic simulated scenarios.

The survey [13] reviews the main streaming techniques and provides a classification of the computational problems and the applications they effectively address, with an emphasis on decentralized scenarios, which are of particular interest in pervasive networks.

Fully decentralized classification and recommendation. [44, 45, 18] describe the first results of ongoing research on fully decentralized, stigmergy-based techniques for nodes of a network to classify interests of users visiting them based only on history information carried by the users themselves. The approach is based on the definition of a statistical model of user behaviour, whose parameters can be estimated from simple statistics over user visits. The performance achieved on data sets from real applications are competitive with those of state-of-art centralized solutions.

[51] proposes fully decentralized techniques to recommend new contacts in a network of mobile phone users, based on similarity between contact list. To this purpose, compact, privacy-preserving, hash-based sketches of users' contact lists are piggybacked over SMS messages.

5.2 Scheduling and resource management

Machine scheduling. A first line of research has concerned the design and analysis of algorithms to schedule jobs on multitasking systems. The focus has been on scheduling policies that optimize performance in terms of user perceived latency. This is typically modelled as the problem of scheduling jobs on one ore more machines, so as to minimize some function of the time spent by jobs in the system. Common measures are the average time spent by each job in the system (flow time), or the derived average stretch. This problem has been studied within the general framework of competitive analysis. A survey of the main results achieved in this direction and an outline of the main techniques used are presented in [35] and have been more recently updated in [54].

The papers [30, 31, 14] show that a randomized variant of the well known Multi Level Feedback (MLF) policy achieves performance close to optimum in expectation in minimizing the average flow time when no information is known about jobs until completion, thus providing a first explanation for the practical effectiveness of a policy which is at the basis of those adopted in most operating systems. [34, 10] study the problem in a different perspective: applying the general framework of *smoothed analysis* initially proposed by Spielman and Teng to the analysis of on-line algorithms, these papers develop sophisticated probabilistics techniques to quantitatively show that MLF's worst case performances are determined by instances that are infrequent and scattered in the input space.

Analysis of the problem in the case of the related average stretch objective is presented in [36, 11]. [33, 9] consider the performance achievable when the processing requirements of tasks are only known approximately. Finally, [32, 15] consider extensions to the case of the weighted flow time objective function.

Team formation and task assignment. The internet has enabled the collaboration of groups at a scale that was unseen before. A key problem for large collaboration groups is to be able to allocate tasks effectively. An effective task assignment method should consider both how *fit* teams are for each job as well as how *fair* the assignment is to team members, in terms that no

one should be overloaded or unfairly singled out. The assignment has to be done automatically or semi-automatically given that it is difficult and time-consuming to keep track of the skills and the workload of each person. Obviously the method to do this assignment must also be computationally efficient.

The papers [20, 22] present a general framework for task-assignment problems. It provides a formal treatment on how to represent teams and tasks and it further proposes alternative functions for measuring the fitness of a team performing a task. Desirable properties of such functions are discussed. The paper [20] then focuses on an important class of task-assignment problems, for which it characterizes the complexity of the problem. Algorithms with provable approximation guarantees, as well as lower bounds are proposed. Finally, the paper presents experimental results that show that the proposed methods are useful in practice in several application scenarios.

The paper [22] extends the results of [20] by considering the computationally challenging, multi-objective scenario in which the performance of a team is also affected by the respective roles and connectedness of the team agents within a social network.

Memory hierarchies. Traditional competitive analysis of well-known paging algorithms (e.g., the Least Recently Used heuristic or derived ones) predicts a performance that is much worse than observed in practice. One reason for this is that a mere combinatorial approach completely dismisses the role played by the phenomenon of locality of reference. [37] proposes an analytical framework that probabilistically models the effects of locality, providing results that are more consistent with practical experience than predicted by standard competitive analysis.

5.3 Network design and management

Research in this area focussed on optimization problems arising in the design and management of wired or wireless networks based on different technologies.

Cooperative mechanisms for network design. [47, 8] propose efficient strategies for the NP-hard problem of designing a network connecting a given set of terminal pairs, so that the cost of the network is minimized. The possibility of an economy of scale (i.e., buying bandwidth rather than renting it when user demand is sufficiently high) is modeled by a concave cost function. The proposed strategies exploit relationships to the problem of network design in the presence of non-cooperative agents.

Bandwidth assignment and channel scheduling in Wireless networks. [25] and [4] present the result of joint research with AT& T research labs on code and power assignment when scheduling the downlink channel in Code Division Multiple Access (CDMA) networks. The problems considered are in general NP-hard. Advanced analysis on a realistic model allows to show that simple heuristics achieve performance close to optimum.

[12] describes results of research performed within EU Integrated Project WINE, aimed at the design and implementation of an adaptation layer, whose purpose is offering a unified interface for the access to wireless LAN segments based on different technologies (e.g., IEEE 802.11 and

BLUETOOTH). The focus of the research described in [12] is on policies to schedule IP packets, so as to provide differentiated Quality of Service to applications, depending on their requirements.

Scheduling and aggregation in Wireless Sensor Networks. Energy comsumption is a major issue in Wireless Sensor Networks. Data aggregation techniques can help strike a balance between energy efficiency and accuracy. [48, 16] attempt a first comprehensive modelling of this setting, with the purpose of investigating the underlying complexity issues. Various models of communication and synchronization are considered and strategies for data collection and delivery under these models are proposed, with the aim of optimizing energy consumption, under latency constraints on data delivery at the sink.

Efficient routing strategies in ATM networks. In ATM (Asynchronous Transfer Mode) networks (and, more recently in Multiprotocol Label Switching ones), packet routing is mostly performed by dedicated hardware. This is made possible by the adoption of a hierarchial routing scheme, whereby packet routing between two end terminals occurs over a virtual channel consisting of one or more virtual paths. Virtual paths in turn consist of one or more physical links, in which routing occurs in a Hw fashion. Designing an ATM implies choosing the set of virtual paths. This choice, among others, affects the size of ATM routing tables, which should be minimized. On the other hand, having more virtual paths reduces the average number of virtual paths a packet traverses in its way to the destination, which in turns affects the latency experienced by the packet. The theoretical foundations of this trade-off is investigated in [27] and [1], which provide a general characterization in terms of parameters describing the underlying physical topology. Efficient routing strategies for specific topologies are proposed in [1].

Call scheduling in all-optical networks. [28, 6] study the problem of scheduling a set of calls over a WDM (Wavelength Division Multiplexing) network, assuming, as is the practical case, that only a limited number of wavelengths is available. The goal is minimizing the overall time needed to schedule calls. The problem consided is in general NP-hard. These contributions propose approximation algorithms for array and ring topology networks. [29, 7] consider the same problem for star and tree topologies.

5.4 Other research interests

Mapping graph partitions on grid topologies. [50] considers the task of decomposing numerical problems into smaller subproblems that are mapped onto the nodes of a parallel computing architecture. Results show that simple partitioning strategies achieve performance close to that of state-of-art partitioning libraries.

Java multicasting. [26] describes the Java implementation of a reliable, multicast version of the FTP protocol. The implementation is tested on an experimental benchmark, emulating a geographical scenario.

5.5 Awards

In September 2000, [56] was awarded the Italian prize for the best PhD thesis in theoretical computer science, by the Italian Chapter of the European Association on Theoretical Computer Science.

5.6 Patent applications

I am co-author of research resulting in the following patent applications:

- Together with F. Delli Priscoli and T. Inzerilli: "Metodo di controllo del traffico in reti a pacchetto Internet". The application was filed by the patent office of University of Rome "La Sapienza" and refers to algorithmic methods for the achievement of predictable QoS in IP based wireless and wired networks, described in [12].

- Together with S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal e A. Vitaletti: "Method for Wireless Downlink Scheduling". This invention refers to methods for QoS management on the downlink of wireless, CDMA based networks described in [25, 4]. The application was filed by AT&T Corporation.

- Together with Aris anagnostopoulos and Aris Gionis: "An optimization framework for query recommendation". The application refers to research described in [19] and was filed by Yahoo! corporation.

6 Service

I am currently local coordinator for Sapienza University of Rome of the following national project, funded by the Italian Ministry of Research:

- PRIN project COGENT: "Aspetti Computazionali e di Teoria dei Giochi in Reti Non-Coordinate".

I am currently responsible for the following projects at Sapienza University of Rome:

- School of Engineering project on "Modelling and visualization of information networks".
- School of Engineering project on "Algorithms for the analysis and management of large scale information systems".

6.1 Research proposal reviewing activity

- **April 2008:** review of a research proposal upon invitation from the "Dutch Organization for Scientific Research" - "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)"

6.2 Program committees

- 15th European Symposium on Algorithms (ESA 2007) - Design and Analysis Track.
- 8th Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP 2007).
- 5th Workshop on Approximation and Online Algorithms (WAOA 2007).
- 2nd ACM International Conference on Web Search and Data Mining (WSDM 2009).
- 34th International Symposium on Mathematical Foundations of Computer Science (MFCS 2009).
- 13th IEEE Conference on Computational Science and Engineering, 2010.
- 9th ACS/IEEE International Conference on Computer Systems and Applications, 2010.
- 21st ACM Conference in Information Knowledge and Management, main session and poster session, 2012.
- ACM Conference on Web Search and Data Mining, main session and doctoral consortium, registration chair.
- IEEE Conference on Computer Science and Engineering, 2012.
- 22nd ACM Conference in Information Knowledge and Management, 2013
- Upcoming: 23rd ACM Conference in Information Knowledge and Management, 2014
- Upcoming: 30th IEEE International Conference on Data Engineering, 2014
- Upcoming: 7th ACM Conference on Web Search and Data Mining, 2014

6.3 Other reviewing activity

International journals. Amongs others: *Elsevier Algorithmica*, *Internet Mathematics*, *ACM Transactions on Knowledge Discovery from Data*, *IEEE Journal on Selected Area in Communications*, *Journal on Computer and Systems Science*, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Algorithms*, *Algorithmica*, *Theoretical Computer Science*, *Discrete and Applied Mathematics*.

International conferences. Among others: *European Symposium on Algorithms (ESA)*, *ACM Symposium on Parallelism in Algorithms and Architectures*, *ACM/SIAM Symposium on Discrete algorithms (SODA)*, *International Colloquium on Automata, Languages and Programming (ICALP)*, *Symposium on Theoretical Aspects of Computer Science (STACS)*, *Conference on Integer Programming and Combinatorial Optimization (IPCO)*, *ACM Symposium on Principles of Distributed Computing (PODC)*, *ACM Symposium on Principles of Database Systems (PODS)*, *Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX)*, *International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*.

6.4 Participation in european research projects

- 1995-1998: EU ESPRIT Long Term Research Project ALCOM-IT;
- 1996-1999: EU ESPRIT Long Term Research Project ALCOM-IT;
- 2000-2003: EU IST Project ALCOM-FT (Algorithms and Complexity in Future Technologies);
- 2000-2001: EU IST Project APPOL (Approximation and On-line Algorithms);
- 2001-2004: EU IST Project APPOL II;
- 2004-2008: EU FET IP Project DELIS: "Dynamically Evolving Large Scale Information Systems".
- 2005-: EU FET IP Project AEOLUS: "Algorithmic Principles for Building Efficient Overlay Computers".
- 2008-: EU FET Project FRONTS: Foundations of Adaptive Networked Societies of Tiny Artefacts.

7 Teaching

I have taught undergraduate and graduate courses at Sapienza University of Rome since the academic year 2000/2001. Among these, undegraduate courses in algorithms and data structures, graduate and undergraduate courses in computer networks.

In the academic year 2009/2010 I am teaching an undegraduate course in foundations of computer science and programming, a graduate course in computer networks and a graduate/PhD course in advanced topics and algorithmic techniques in computer networking.

8 References

Journals

[1] L. Becchetti, P. Bertolazzi, C. Gaibisso, and G. Gambosi. On the design of efficient ATM routing schemes. *Theoretical Computer Science*, 270(1–2):341–359, 2002.

[2] L. Becchetti, C. Castillo, D. Donato, S. Leonardi, and R. Baeza-Yates. Link analysis for web spam detection. *ACM Trans. on the Web*, 2(1):1–42, 2007.

[3] L. Becchetti, U. Colesanti, A. Marchetti-Spaccamela, and A. Vitaletti. Fully decentralized recommendations in pervasive systems: Models and experimental analysis. *International Journ Engineering Intelligent Systems Electrical Engineering Communications*, 20(3):161, 2012.

[4] L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal, and A. Vitaletti. Parallel scheduling problems in next generation wireless networks. *Elsevier NETWORKS*, 45(1):9 – 22, 2005.

[5] L. Becchetti, L. Filipponi, and A. Vitaletti. Privacy support in people-centric sensing. *Journal of Communications*, 7(8):606–621, 2012.

[6] L. Becchetti, M. Di Ianni, and A. Marchetti-Spaccamela. Approximation algorithms for routing and call scheduling in all-optical chains and rings. *Theoretical Computer Science*, 287(2):429–448, 2002.

[7] L. Becchetti, M. Di Ianni, and A. Marchetti-Spaccamela. Approximating call-scheduling makespan in all-optical networks. *Journal of Discrete Algorithms*, 2(4):501–515, 2004.

[8] L. Becchetti, J. Könemann, S. Leonardi, and M. Pál. Sharing the cost more efficiently: Improved approximation for multicommodity rent-or-buy. *ACM Trans. Algorithms*, 3(2), 2007.

[9] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, and Kirk Pruhs. Semi-clairvoyant scheduling. *Theoretical Computer Science*, 324:325 – 335, 2004.

[10] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, G. Schaefer, and T. Vredeveld. Average case and smoothed competitive analysis of the multi-level feedback algorithm. *Mathematics of Operations Research*, 31:85–108, 2006.

[11] L. Becchetti, S. Leonardi, and S. Muthukrishnan. Average stretch without migration. *Journal of Computer and System Sciences*, 68:80–95, 2004.

[12] L. Becchetti, F. Delli Priscoli, T. Inzerilli, P. Mähönen, and L. Muñoz. Enhancing IP service provision over heterogenous wireless networks: A path toward 4G. *IEEE Communications Magazine*, 39(8):74–81, 2001.

[13] Luca Becchetti, Ioannis Chatzigiannakis, and Yiannis Giannakopoulos. Streaming techniques and data aggregation in networks of tiny artifacts. *Computer Science Review, Elsevier*, 2010. Final version accepted for publication.

[14] Luca Becchetti and Stefano Leonardi. Nonclairvoyant scheduling to minimize the total flow time on single and parallel machines. *J. ACM*, 51(4):517–539, 2004.

[15] Luca Becchetti, Stefano Leonardi, Alberto Marchetti-Spaccamela, and Kirk Pruhs. Online weighted flow time and deadline scheduling. *J. Discrete Algorithms*, 4(3):339–352, 2006.

[16] Luca Becchetti, Alberto Marchetti-Spaccamela, Andrea Vitaletti, Peter Korteweg, Martin Skutella, and Leen Stougie. Latency-constrained aggregation in sensor networks. *ACM Trans. Algorithms*, 6(1):1–20, 2009.

[17] Becchetti L., Boldi P., Castillo C., and Gionis A. Efficient algorithms for large-scale local triangle counting. *ACM Transactions on Knowledge Discovery from Data*, 4:13:1–13:28, 2010.

[18] Becchetti L., Colesanti U., Marchetti-Spaccamela A., and Vitaletti A. Recommending items in pervasive scenarios: models and experimental analysis. *Knowledge and Information Systems*, pages 1–24, 2010.

Conferences and Workshops

[19] Anagnostopoulos A., Becchetti L., Castillo C., and Gionis A. An optimization framework for query recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 161–170. ACM, 2010.

[20] Anagnostopoulos A., Becchetti L., Castillo C., Gionis A., and Leonardi S. Power in unity: forming teams in large-scale community systems. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 599–608. ACM, 2010.

[21] Anagnostopoulos A., Becchetti L., Leonardi S., Mele I., and Sankowski P. Stochastic query covering. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 725–734. ACM, 2011. Best paper award in poster session.

[22] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi. Online team formation in social networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 839–848. ACM, 2012.

[23] L. Becchetti, L. Bergamini, F. Ficarola, and A. Vitaletti. Population protocols on real social networks. In *Proceedings of the Fifth Workshop on Social Network Systems*, page 15. ACM, 2012.

[24] L. Becchetti, P. Boldi, C. Castillo, and A. Gionis. Efficient semi-streaming algorithms for local triangle counting in massive graphs. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2008)*, pages 16–24. ACM Press, 2008.

[25] L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal, and A. Vitaletti. Parallel scheduling problems in next generation wireless networks. In *Proceedings of the Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 238–247, 2002.

[26] L. Becchetti, M. Draoli, C. Gaibisso, and F. Lombardi. A Java Implementation of a Reliable Multicast File Transfer Protocol: Design and Evaluation. In *Proceedings of the IASTED International Conference on INTERNET AND MULTIMEDIA SYSTEMS AND APPLICATIONS*, pages 163–168, 2001.

[27] L. Becchetti and C. Gaibisso. Lower bounds for the virtual path layout problem in ATM networks. In *Proceedings of Theory and Practice of Informatics, Seminar on Current Trends in Theory and Practice of Informatics, Lectures Notes in Computer Science*, volume 24, pages 375–382, 1997.

[28] L. Becchetti, M. Di Ianni, and A. Marchetti-Spaccamela. Approximation algorithms for routing and call scheduling in all-optical chains and rings. In *Proceedings of FSTTCS: Foundations of Software Technology and Theoretical Computer Science, Lectures Notes in Computer Science*, pages 201–212, 1999.

[29] L. Becchetti, M. Di Ianni, and A. Marchetti-Spaccamela. Approximating call-scheduling makespan in all-optical networks. In *Proceedings of WG: International Workshop on Graph-Theoretic Concepts in Computer Science, Lectures Notes in Computer Science*, pages 13–22, 2000.

[30] L. Becchetti and S. Leonardi. Non-clairvoyant scheduling to minimize the average flow time on single and parallel machines. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 94–103, 2001.

[31] L. Becchetti and S. Leonardi. Non-clairvoyant scheduling to minimize the average flow time on single and parallel machines (invited paper). In *Proceedings of the 2th International Workshop on Approximation and Randomization Algorithms in Communication Networks (ARACNE 2001)*, pages 1–6, 2001.

[32] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, and K. Pruhs. Online weighted flow time and deadline scheduling. In *Proceedings of RANDOM-APPROX, Lectures Notes in Computer Science*, pages 36–47, 2001.

[33] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, and Kirk Pruhs. Semi-clairvoyant scheduling. In *Proceedings of the 11th European Symp. on Algorithms (ESA), Lectures Notes in Computer Science*, pages 67–77, 2003.

[34] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, G. Schaefer, and T. Vredeveld. Average case and smoothed competitive analysis of the multi-level feedback algorithm. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 462–471, 2003.

[35] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, and G. Schäfer. Scheduling to minimize flow time metrics. In *Proceedings of the IEEE International Parallel and Distributed Symposium*, page 223, 2003.

[36] L. Becchetti, S. Leonardi, and S. Muthukrishnan. Scheduling to minimize average stretch without migration. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 548–557, 2000.

[37] Luca Becchetti. Modeling locality: A probabilistic analysis of LRU and FWF. In *Proceedings of ESA 2004, 12th Annual European Symposium on Algorithms*, volume 3221 of *Lecture Notes in Computer Science*, pages 98–109. Springer, 2004.

[38] Luca Becchetti, Lorenzo Bergamini, Francesco Ficarola, Francesco Salvatore, and Andrea Vitaletti. First experiences with the implementation and evaluation of population protocols on physical devices. In *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*, pages 335–342. IEEE, 2012.

[39] Luca Becchetti, Lorenzo Bergamini, Francesco Ficarola, and Andrea Vitaletti. Population protocols on real social networks. In *Proceedings of the Fifth Workshop on Social Network Systems*, SNS '12, pages 15:1–15:2. ACM, 2012.

[40] Luca Becchetti and Carlos Castillo. The distribution of pagerank follows a power-law only for particular values of the damping factor. In *Proceedings of the 15th international conference on World Wide Web, WWW 2006*, pages 941–942. ACM, 2006.

[41] Luca Becchetti, Carlos Castillo, Debora Donato, and Adriano Fazzone. A comparison of sampling techniques for web characterization. In *Prceedings of the Workshop on Link Analysis (LinkKDD)*. ACM, 2006.

[42] Luca Becchetti, Carlos Castillo, Debora Donato, Stefano Leonardi, and Ricardo Baeza-Yates. Using rank propagation and probabilistic counting for link-based spam detection. In *Proceedings of the Workshop on Web Mining and Web Usage Analysis (WebKDD)*. ACM Press, 2006.

[43] Luca Becchetti, Carlos Castillo, Debora Donato, Stefano Leonardi, and Ricardo A. Baeza-Yates. Link-based characterization and detection of web spam. In *Proceedings of AIRWeb: Workshop on Adversarial Information Retrieval on the Web*, pages 1–8, 2006.

[44] Luca Becchetti, Ugo Colesanti, Alberto Marchetti-Spaccamela, and Andrea Vitaletti. Self-adaptive recommendation systems: Models and experimental analysis. In *Proceedings of the Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems*. IEEE, 2008. final version accepted for publication in poster session.

[45] Luca Becchetti, Ugo Colesanti, Alberto Marchetti-Spaccamela, and Andrea Vitaletti. Fully decentralized recommendations in pervasive systems: Models and experimental analysis. In *Proceedings of the First International Workshop on Data Warehousing and Knowledge Discovery from Sensors and Streams*, 2009.

[46] Luca Becchetti, Luca Filipponi, and Andrea Vitaletti. Opportunistic privacy preserving environmental noise pollution monitoring. In *Proceedings of the International Workshop on Sensing for App Phones*. Dartmouth University, 2010. ACM-Sensys workshop, procs. available at http://sensorlab.cs.dartmouth.edu/phonesense/proceeding.pdf.

[47] Luca Becchetti, Jochen Könemann, Stefano Leonardi, and Martin Pál. Sharing the cost more efficiently: improved approximation for multicommodity rent-or-buy. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, pages 375–384. SIAM, 2005.

[48] Luca Becchetti, Peter Korteweg, Alberto Marchetti-Spaccamela, Martin Skutella, Leen Stougie, and Andrea Vitaletti. Latency constrained aggregation in sensor networks. In *Proceedings of ESA 2006, 14th Annual European Symposium on Algorithms*, volume 4168 of *Lecture Notes in Computer Science*, pages 88–99. Springer, 2006.

[49] Luca Becchetti and Elias Koutsoupias. Competitive analysis of aggregate max in windowed streaming. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming*, 2009.

[50] F. D'Amore, L. Becchetti, S. L. Bezrukov, and A. Marchetti-Spaccamela. On the embedding of refinements of 2-dimensional grids. In *Proceedings of Europar '97, Lectures Notes in Computer Science*, pages 950–957, 1997.

[51] Baglioni E., Becchetti L., Bergamini L., Colesanti U., Filipponi L., Vitaletti A., and Persiano G. A lightweight privacy preserving sms-based recommendation system for mobile users. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 191–198. ACM, 2010.

[52] Becchetti L., Bordino I., Leonardi S., and Rosen A. Fully decentralized computation of aggregates over data streams. In *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, pages 1–9. ACM, 2010. Best paper award. Also published on-line in SIG-KDD explorations.

Book chapters

[53] *Optimisation combinatoire*, chapter Algorithmes on-line. Hermes, 2005.

[54] *Encyclopedia of Algorithms*, chapter Flowtime Minimization. Springer, 2008. Invited contribution.

On-line magazines

[55] Carlos Castillo, Debora Donato, Luca Becchetti, Paolo Boldi, Stefano Leonardi, Massimo Santini, and Sebastiano Vigna. A reference collection for web spam. *ACM SIGIR Forum*, 40(2):11–24, 2006.

Technical reports

[56] L. Becchetti. Efficient resource management high bandwidth networks. PhD thesis, School of Engineering, Sapienza University of Rome, 1999.

[57] Luca Becchetti, Vincenzo Bonifaci, Michael Dirnberger, Andreas Karrenbauer, and Kurt Mehlhorn. Physarum can compute shortest paths: Convergence proofs and complexity bounds. To appear in the proceedings of the 40th International Colloquium on Automata, Languages and Programming (ICALP 2013), 2013.