



POLO CHAU

Legal name:
Duen Horng Chau

Machine Learning Department
School of Computer Science
Carnegie Mellon University

dchau@cs.cmu.edu
www.cs.cmu.edu/~dchau

EDUCATION

- Aug, 2012 (expected) **Ph.D. Machine Learning** Carnegie Mellon University
Understanding and Interacting with Massive Graphs: Bridging Data Mining and Human-Computer Interaction
Committee: [Christos Faloutsos](#) (Chair, Advisor), [Jason Hong](#) (Co-advisor), [Niki Kittur](#), [Jiawei Han](#) (UIUC)
- 2010 **M.S. Machine Learning** Carnegie Mellon University
- 2005 **M.S. Human-Computer Interaction** Carnegie Mellon University
- 2004 **B.Eng. Information Engineering (Honors)** The Chinese University of Hong Kong
Online Auction Fraud Detection
Advisor: [Prof. Soung-Chang Liew](#)
- 2000 **Diocesan Boys' School** (high school) Hong Kong

RESEARCH INTERESTS

My research *bridges* **Data Mining** and **Human-Computer Interaction** to synthesize *systems* and *tools* that help people understand and interact with **Big Data**. My thesis focuses on **massive networks** with *billions* of nodes and edges.

I blend techniques from **machine learning** (Belief Propagation), **data mining** (anomaly detection), **visualization** and **user interaction**. Notable projects:

- Apolo** mixed-initiative system (ML+vis) for making sense of large network data (watch a [demo](#));
Apolo helps users incrementally find relevant subgraphs to explore, avoids overwhelming the user
- Polonium** patent-pending malware detection technology (with Symantec);
infers file reputation over **37 billion** machine-file relationships
- NetProbe** auction fraud detection (eBay) that fingers bad guys by identifying their suspicious transactions;
appeared in Wall Street Journal, USA Today, and more
- PEGASUS** award-winning project that creates graph algorithms for massive graph data

ACADEMIC HONORS & AWARDS

- 2010 **Open Source Software World Challenge, Silver Award (3rd place)**, for the [PEGASUS project](#) which mines billion-node graphs. U Kang, Duen Horng Chau, Christos Faloutsos
- 2009 - 2010 **Symantec Research Labs Graduate Fellowship**, covers full tuition and stipend. Re-selected, as one of the only three graduate students worldwide to receive the award.
- 2009 **Yahoo! Key Scientific Challenges Award**, in Information Retrieval, Algorithms, and Data Mining.
- 2008 - 2009 **Symantec Research Labs Graduate Fellowship**, covers full tuition and stipend.
- 2008 **Winner**, Symantec research competition (internal).
- 2006 **Best Presentation Award, 3rd Place**, PKDD'06.
- 2001, 2002, 2003 Dean's List
- 2002, 2003 Nominated for university scholarships

EMPLOYMENT

- Summer-Fall 2011 **Google, Ads Backend Team** *Mountain View, CA*
Ph.D. Software Engineer Intern Mentor: Dr. Arun Swami
- Fall 2009 **Symantec Research Labs** *Los Angeles, CA*
Ph.D. Research Intern Mentor: Mr. Carey Nachenberg
- Summer 2008 **Symantec Research Labs** *Los Angeles, CA*
Ph.D. Research Intern Mentor: Mr. Darren Shou
- 2005 - 2007 **Carnegie Mellon University, Human-Computer Interaction Institute** *Pittsburgh, PA*
Research Associate Supervisor: Prof. Brad Myers
- Summer 2003 **The Chinese University of Hong Kong, Lightwave Communications Lab** *Hong Kong*
Undergraduate Research Assistant
- 2002 - 2007 **Elfware Company** Hong Kong
System Consultant

TEACHING

- Guest lectures **Web Search** for *Science of the Web (15-396)*.
Instructor: Prof. Luis von Ahn, Brendan Meeder. Fall 2011. Carnegie Mellon University.
- Making Sense of Large Networks** for *Sensemaking: Cognitive, Social, and Technical Perspectives (05-899)*.
Instructor: Prof. Niki Kittur. Spring 2011. Carnegie Mellon University.
- Teaching Assistant **Science of the Web (15-396)**.
Instructor: Prof. Luis von Ahn, Brendan Meeder. Fall 2011. Carnegie Mellon University.

PUBLICATIONS (Google h-index: 12)

Conference Papers

PEGASUS: Mining Billion-Scale Graphs in the Cloud. U Kang, Duen Horng Chau, and Christos Faloutsos. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2012*. Mar 25-30, 2012. Kyoto, Japan

Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms.

Danai Koutra, Tai-You Ke, U Kang, Duen Horng (Polo) Chau, Hsing-Kuo Kenneth Pao, and Christos Faloutsos. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD) 2011*. Sept 5-9, 2011. Athens, Greece. [20% acceptance rate]

[PDF](#)

Mining Large Graphs: Algorithms, Inference, and Discoveries. U Kang, Duen Horng (Polo) Chau, Christos Faloutsos. *IEEE International Conference of Data Engineering (ICDE) 2011*. April 11-16. Hannover, Germany. [20% acceptance rate]

[PDF](#)

Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning. Duen Horng (Polo) Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos. *ACM Conference on Human Factors in Computing Systems (CHI) 2011*. May 7-12, 2011. Vancouver, BC, Canada. [26% acceptance rate]

[PDF](#)
[YouTube](#)



Polonium: Tera-Scale Graph Mining and Inference for Malware Detection. Duen Horng (Polo) Chau, Carey Nachenberg, Jeffrey Wilhelm, Adam Wright, Christos Faloutsos. *SIAM International Conference on Data Mining (SDM) 2011*. April 28-30, 2011. Mesa, Arizona. [25% acceptance rate]

[PDF](#)

On the Vulnerability of Large Graphs. Hanghang Tong, B. Aditya Prakash, Charalampos Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, Duen Horng (Polo) Chau. *IEEE International Conference on Data Mining (ICDM) 2010*. Dec 14-17, 2010. Sydney, Australia. [19% acceptance rate]

[PDF](#)
[Code](#)

What to Do When Search Fails: Finding Information by Association. Duen Horng (Polo) Chau, Brad Myers, and Andrew Faulring. *ACM Conference on Human Factors in Computing Systems (CHI) 2008*. April 5-10, 2008. Florence, Italy. New York: ACM Press, Pages 999-1008. [22% acceptance rate]

[PDF](#)
[YouTube](#)



The eBay Graph: How Do Online Auction Users Interact? Yordanos Beyene, Michalis Faloutsos, Duen Horng (Polo) Chau, Christos Faloutsos. *Computer Communications Workshops, 2008. INFOCOM*. April 13-18, 2008. Phoenix, AZ.

[PDF](#)

NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. Shashank Pandit, Duen Horng (Polo) Chau, Samuel Wang, Christos Faloutsos. *International Conference on World Wide Web (WWW) 2007*. May 8-12, 2007. Banff, Alberta, Canada. Pages 201-210. [15% acceptance rate]

[PDF](#)

Eyes on the Road, Hands on the Wheel: Thumb-based Interaction Techniques for Input on Steering Wheels. Ivan E. Gonzalez, Jacob O. Wobbrock, Duen Horng (Polo) Chau, Andrew Faulring, Brad A. Myers. *Graphics Interface (GI) 2007*. May 28-30, 2007. Montreal, Quebec, Canada. Pages 95-102. [48% acceptance rate]

[PDF](#)

Demonstrating the Viability of Automatically Generated User Interfaces. Jeffrey Nichols, Duen Horng (Polo) Chau, Brad A. Myers. *ACM Conference on Human Factors in Computing Systems (CHI) 2007*. April 28-May 3, 2007. San Jose, CA. Pages 1283-1292. [25% acceptance rate]

[PDF](#)

An Alternative to Push, Press, and Tap-tap-tap: Gesturing on an Isometric

[PDF](#)

Joystick for Mobile Phone Text Entry. Jacob O. Wobbrock, Duen Horng (Polo) Chau and Brad A. Myers. *ACM Conference on Human Factors in Computing Systems (CHI) 2007*. April 28-May 3, 2007. San Jose, CA. Pages 667-676. [25% acceptance rate]

Huddle: Automatically Generating Interfaces for Systems of Multiple Connected Appliances. Jeffrey Nichols, Brandon Rothrock, Duen Horng (Polo) Chau, Brad A. Myers. *ACM Symposium on User Interface Software and Technology (UIST) 2006*. October 15-18, 2006. Montreux, Switzerland. Pages 279-288. [23% acceptance rate]

[PDF](#)

In-stroke Word Completion. Jacob O. Wobbrock, Brad A. Myers, and Duen Horng (Polo) Chau. *ACM Symposium on User Interface Software and Technology (UIST) 2006*. October 15-18, 2006. Montreux, Switzerland. Pages 333-336. [23% acceptance rate]

[PDF](#)

Detecting Fraudulent Personalities in Networks of Online Auctioneers. Duen Horng (Polo) Chau, Shashank Pandit, and Christos Faloutsos. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD) 2006*. Sept 18-22, 2006. Berlin, Germany. Pages 103-114. [15% acceptance rate]

[PDF](#)

Best Presentation Award, 3rd place

A Linguistic Analysis of How People Describe Software Problems. Andrew J. Ko, Brad A. Myers, and Duen Horng (Polo) Chau. *Proceedings of VL/HCC 2006*. September 4-8, 2006, Brighton, UK. Pages 127-134. [28% acceptance rate]

[PDF](#)

Answering Why and Why Not Questions in User Interfaces. Brad Myers, David A. Weitzman, Andrew J. Ko, and Duen Horng (Polo) Chau. *ACM Conference on Human Factors in Computing Systems (CHI) 2006*. April 22-27, 2006. Montreal, Canada. Pages 397-406. [23% acceptance rate]

[PDF](#)

Invited Articles

Catching Bad Guys with Graph Mining. Polo Chau. *Crossroads: The ACM Magazine for Students - The Fate of Money*. Volume 17 Issue 3, Spring 2011. Pages 16-18.

[PDF](#)

Case study on fraud detection using social-network analysis. Duen Horng (Polo) Chau and Christos Faloutsos. *Encyclopedia of Social Networks and Mining*. In preparation.

Workshop Papers

Polonium: Tera-Scale Graph Mining for Malware Detection. Duen Horng (Polo) Chau, Carey Nachenberg, Jeffrey Wilhelm, Adam Wright, Christos Faloutsos. *The 2nd Workshop on Large-scale Data Mining: Theory and Applications (LDMTA 2010)*. July 25, 2010. Washington, DC.

IBM Student Travel Fellowship

Inference of Beliefs on Billion-Scale Graphs. U Kang, Duen Horng (Polo) Chau, Christos Faloutsos. *The 2nd Workshop on Large-scale Data Mining: Theory and Applications (LDMTA 2010)*. July 25-28, 2010. Washington, DC.

[PDF](#)

Supporting Ad Hoc Sensemaking: Integrating Cognitive, HCI, and Data Mining Approaches. Aniket Kittur, Duen Horng (Polo) Chau, Christos Faloutsos, Jason I. Hong. *Sensemaking Workshop at CHI 2009*. April 4-5, 2009. Boston, MA.

Feldspar: A System for Finding Information by Association. Duen Horng (Polo) Chau, Brad Myers, and Andrew Faulring. *PIM 2008: CHI 2008 Workshop on Personal Information Management*. April 5-6, 2008. Florence, Italy.

[PDF](#)

Fraud Detection in Electronic Auction. Duen Horng (Polo) Chau and Christos Faloutsos. *Proceedings of EWMF'05: European Web Mining Forum, at ECML/PKDD'05*. October 3-7, 2005. Porto, Portugal.

[PDF](#)

Poster Papers **TopicViz: Semantic Navigation of Document Collections.** Jacob Eisenstein, Duen Horng (Polo) Chau, Aniket Kittur, Eric P. Xing. *Extended Abstracts, CHI 2012*. May 5-10, 2012. Austin, TX, USA.

Parallel Crawling for Online Social Networks. Duen Horng (Polo) Chau, Shashank Pandit, Samuel Wang, and Christos Faloutsos. *International Conference on World Wide Web (WWW) 2007*. May 8-12, 2007. Banff, Alberta, Canada. Pages 201-210.

Integrating Isometric Joysticks into Mobile Phones for Text Entry. Duen Horng (Polo) Chau, Jacob O. Wobbrock, Brad A. Myers, Brandon Rothrock. *Extended Abstracts, CHI 2006*. April 22-27, 2006. Montreal, Canada. Pages 640-645.

[PDF](#)

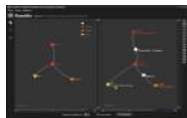
Demos **OPAvion: Mining and visualization in large graphs.** Leman Akoglu, Duen Horng (Polo) Chau, U Kang, Danai Koutra, Christos Faloutsos. *ACM SIGMOD Conference 2012*. May 20-24, 2012. Scottsdale, Arizona, USA

Apolo: Interactive Large Graph Sensemaking by Combining Machine Learning and Visualization. Duen Horng (Polo) Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2011*. Aug 21-21, 2011. San Diego, California, USA.

[PDF](#)

SHIFTR: A Fast and Scalable System for Ad Hoc Sensemaking of Large Graphs. Duen Horng (Polo) Chau, Aniket Kittur, Christos Faloutsos, Hanghang Tong, Jason I. Hong. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2009*. June 28-July 1, 2009. Paris, France.

[PDF](#)



GRAPHITE: A Visual Query System for Large Graphs. Duen Horng (Polo) Chau, Christos Faloutsos, Hanghang Tong, Jason I. Hong, Brian Gallagher, Tina Eliassi-Rad. *International Conference on Data Mining (ICDM) 2008*. Dec 15-19, 2008. Pisa, Italy.

[PDF](#)
[YouTube](#)

Video **SHIFTR: a user-directed, link-based system for ad hoc sensemaking of large heterogeneous data collections.** Duen Horng (Polo) Chau, Aniket Kittur, Christos Faloutsos, Jason I. Hong. *Extended Abstracts of CHI 2009*. Apr 4-9, 2009. Boston, MA.

DESIGN AWARDS & LEADERSHIP

2006 **Winner**, Carnegie Mellon ID card design contest. My design has been in use since then.

Winner, *Graduate Student Assembly* logo design contest. Carnegie Mellon University.

Winner, *Language Technologies Institute* T-shirt design contest. Carnegie Mellon University.

2005 **Design lead**, HCI Masters capstone project. Client: *Office of Naval Research*.

2004 **Winner**, webpage design contest for the *New Asia College* homepage. The Chinese University of Hong Kong.

Lead graphics designer, The Chinese University of Hong Kong 40th Anniversary Concert.

1999 - 2000 **Art Director** of school yearbook. Diocesan Boys' School (Hong Kong).

TALKS

May 13, 2011 **Mining Massive Graphs: Visualization and Anomaly Detection**
Google. *Mountain View, CA, USA*

May 10, 2011 **Mining Massive Graphs: Visualization and Anomaly Detection**
Simon Fraser University. *Vancouver, Canada* Host: Prof. Jian Pei

Feb 16, 2007 **Detecting Fraudulent Personalities in Networks of Online Auctioneers**
eBay. *Palo Alto, CA, USA*

PATENT APPLICATIONS

Inferring File and Website Reputations by Belief Propagation Leveraging Machine Reputation. Adam Wright, Duen Horng Chau. Filed 22 Feb 2010.

PRESS

"Fraudsters on Internet Auction Sites Often Leave Trail Leading to Crimes", Lee Gomes, **The Wall Street Journal**, December 6, 2006.

"Researchers: Software Can Nip eBay Fraud in The Bud", Bob Sullivan, **MSNBC**, December 8, 2006.

KQV Radio live interview, Host: Frank Gottlieb, December 10, 2006.

"Researchers Developing Anti-Fraud Tool", Joe Mandak, **The Associated Press**, December 11, 2006. Also appeared in over 100 newspapers around the world, including **USA Today**, **Los Angeles Times**, **The Washington Post**.

"CMU Develops Program To Online Auction Fraud", Keith Jones, **KDKA-TV News**, originally aired December 12, 2006.

BizRadio live interview, Host: Brent Clenton, December 13, 2006

"Online Auction Frauds Not As Clever As They Think, Researchers Say", Alpha Doggs, **Network World**, December 5, 2006.

GRANTS

\$35 million DARPA proposal on *Anomaly Detection at Multiple Scales (ADAMS)*. CMU funded ~\$500,000.
Co-authored the winning proposal, represented CMU and led research efforts.

Research proposals led to two full years of research funding from Symantec. Awarded ~\$140,000.

Helped in creating an NSF proposal *HCI meets Data Mining — Attention Routing, Visualization and Sense-making of Large Graphs*. \$500,000

Helped in creating an NEC proposal for *University Awards from NEC Labs Data Management*. \$75,000

ADVISING

2010 **Sahil Jolly**, ECE undergraduate student. Design, implementation and evaluation of large scale graph mining algorithms.

- 2009 **Neha Reddy**, M.S. Information Networking -- now Software Engineer at Cvent.
- 2009 **Karthik Thiyagarajan**, M.S. Information Networking -- now Software Engineer at QuantiFind.
- 2006 - 2007 **Samuel Wang**, CS undergraduate student now CS Ph.D. student at MIT.

ACADEMIC COMMUNITY ACTIVITIES

- Program Committee
 - NIPS 2011 workshop on parallel and large-scale machine learning (**Big Learning**)
 - CSCW 2012 workshop on **Personal Information Management (PIM)**
 - ASIST 2009 workshop on **Personal Information Management (PIM)**
- Organizer
 - Machine Learning Lunch Seminar (2008-2011)**
Machine Learning Department, Carnegie Mellon University
 - The First Student Research Symposium (2009)**
Machine Learning Department, Carnegie Mellon University
- Journal Reviewer
 - IEEE Intelligent Systems (2012)
 - Parallel Computing (2011)
 - IBM Systems Journal (2007)
- Conference Reviewer
 - CHI 2004-2012, UIST 2007-2010, IUI 2012, WWW 2008-2010, ECML/PKDD 2008-2009, VLDB 2008-2009
- Member
 - Association for Computing Machinery (ACM). Since 2006.
 - The Institute of Electrical and Electronics Engineers (IEEE). Since 2006.
 - Machine Learning Department website committee

UNIVERSITY SERVICE

- Mascot Task Force* committee member. Carnegie Mellon University. 2006.
- The Chinese University of Hong Kong 40th Anniversary Concert* committee member. 2004.

COURSEWORK

- Machine Learning & Data Mining
 - Machine learning Statistical machine learning Applied Bayesian methods Multimedia databases and data mining Probabilistic graphical models Algorithms in the real world Intermediate statistics Art that learns
- HCI
 - Software architecture for user interface Programming usable interfaces Computer-mediated communication Human factors Interface and Interaction Design
- Information Engineering
 - Database programming Software engineering Simulation and statistical analysis Network software design and programming Video and image coding and processing Cryptography, security and eCommerce Advanced engineering math

SKILLS

HCI	Contextual inquiry GOMS	Persona	Think-aloud protocol Focus groups	Cognitive walkthrough Rapid paper prototyping	Heuristic evaluation		
Programming	Java ASP	C++ XML	C HTML	SQL	VB Assembly JavaScript ActionScript	CSS	PHP
Design	Photoshop	Illustrator	Premiere	InDesign	Flash	Lightwave 3D	3D Studio MAX
Music	Pianist (20+ years) music group member	Cellist (20+ years) Choir member	High school string orchestra conductor	Chamber			
Languages	English	Mandarin	Cantonese (native)				

REFERENCES

Christos Faloutsos

Professor, Computer Science Department, Carnegie Mellon University

Jason I. Hong

Associate Professor, Human-Computer Interaction Institute, Carnegie Mellon University

Aniket Kittur

Assistant Professor, Human-Computer Interaction Institute, Carnegie Mellon University

Brad Myers

Professor, Human-Computer Interaction Institute, Carnegie Mellon University

Jacob O. Wobbrock

Associate Professor, The Information School, University of Washington

Carey Nachenberg

Vice President and Symantec Fellow, Symantec Corporation
Adjunct Professor of Computer Science, UCLA Computer Science Department

Dr. Arun Swami

Member of Technical Staff, Google

Last modified: April 13, 2012