

# Andy TWIGG

✉ [andy.twigg@gmail.com](mailto:andy.twigg@gmail.com) | ☎ 65024504007 | [in andytwigg](https://www.linkedin.com/in/andytwigg) | [andytwigg](https://github.com/andytwigg) | 📍 Belmont, CA

I'm a computer scientist with expertise in large-scale data processing and ML. I currently lead ML efforts at Lacework, a cloud security company treating security as a big data problem. I have been an entrepreneur (founder/CTO of 2 acquired startups) and an academic (PhD Cambridge, and I was a Fellow in Computer Science at Oxford University 2008-13).

I'm excited about hands-on technical leadership at the intersection of large-scale data and ML, in particular, using RL and synthetic data with LLMs.

## EXPERIENCE

---

### Lacework — *Distinguished Engineer* 2020 - PRESENT

- Lead core 'polygraph' technology: large-scale data and ML pipelines. Processing 5B msgs/hr from agent and cloud logs
- Rewrote core processing from **go** to **pyspark**, EC2 savings of ~\$2m/yr, ~10x reduction in SEVs
- Created and led implementation of inductive GNN-based models using **pytorch-geometric** and streaming density estimation
- Company raised \$1.8B at \$8.3B valuation, acquired by Fortinet

### Milliways Ventures — *Entrepreneur-in-residence* 2018 - 2020

- Exploring ideas around trading with **RL**
- Collected several TBs of **L3** (raw message) data from a large crypto exchange, wrote a simulator with better queue estimation.
- Trained models using **A3C**, **PPO**, etc.

<https://github.com/andytwigg/deeptrade>

<https://medium.com/@andytwigg/learning-to-trade-with-deep-rl-666ed6bbd921>

### C9 (acquired by Insidesales.com) — *CTO/Chief Scientist* 2014 - 2017

- Built one of the first systems using ML to improve sales team efficiency
- Implemented bottom-up forecasting, opportunity scoring by modeling as MDPs
- Acquired in 2015 by Insidesales.com

### Featurestream.io — *Founder* 2013 - 2013

- Built **streaming random forest** (based on Hoeffding trees) on **Spark streaming**
- Hosted as a scalable API on AWS, did early POCs

<https://github.com/featurestream/featurestream-core/>

<https://medium.com/@andytwigg/featurestream-io-random-forests-6992b03b521>

### Oxford University (St Johns College) — *Fellow in Computer Science* 2008 - 2013

- JRF, elected by international open competition
- Took a sabbatical to found Acunu

Followup work on cache-oblivious indexing algorithms: <https://arxiv.org/abs/1707.08186>

### Acunu (acquired) — *Founder, CTO* 2009 - 2012

- Big data startup, built a streaming analytics system on **Cassandra**
- Invented **Stratified B-tree**, a versioned index optimized for SSDs and big data ([Linkedin techtalk](#))
- Hosted as a scalable API on AWS, did early POCs
- Created London Big Data meetup group. Raised \$9m, acquired

### Microsoft Research (Cambridge) and Technicolor Research (Paris) — *Researcher* 2006 - 2008

- [Algorithms for P2P streaming](#) with optimal throughput/latency tradeoffs

## EDUCATION

---

### Cambridge University (King's College) — *PhD, Computer Science* - 2006

Thesis: [Approximate graph routing with failures](#). Nominated for BCS Best Dissertation Award.

### Warwick University — *BSc, Computer Science* - 2002

Top 1st

## INTERESTS

---

I enjoy DIY, drumming, golf. I rowed for Cambridge Lightweights and King's College men's 1st VIII

## SELECTED PUBLICATIONS

---

- Persistent Cache-oblivious Streaming Indexes*, arxiv, abs/1707.08186, 2017
- Locality-preserving allocations problems and coloured bin packing* with E Xavier., J. Theoretical CS, 2015
- Stratified B-trees and versioned dictionaries* Twigg et al, HotStorage 2011
- Constrained-path labellings on graphs of bounded clique-width* with B Courcelle, Theory Comput. Syst., 2010
- Epidemic live streaming: optimal performance trade-offs* Bonald et al, SIGMETRICS, 2008.
- Worst-case time decremental connectivity and k-edge witness problems* ArXiv,abs/0810.5477, 2008
- Connectivity checking in 3-connected planar graphs with obstacles* with Courcelle et al., Notes in Disc Math, 2008
- Rate-optimal schemes for peer-to-peer live streaming* with L. Massoulie, J Perf Eval , 65(11-12):804–822, 2008
- Randomized decentralized broadcasting algorithms* with Massoulie et al, INFOCOM 2007
- Forbidden-set labelling on graphs* with Courcelle et al. PODC (LOCALITY) 2007
- Compact forbidden-set routing* with B Courcelle, STACS 2007
- Complexity of fixed point models of trust in distributed networks* with K Krukow, J Theoretical CS 2007
- Compact forbidden-set routing (PhD Thesis)*, Cambridge TR UCAM-CL-TR-678, 2006
- Provably optimal decentralized broadcasting algorithms* with Massoulie et al, MSR-TR-2006-105