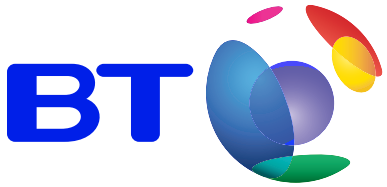
BT Research and MSc projects

Keith Briggs Keith.Briggs@bt.com

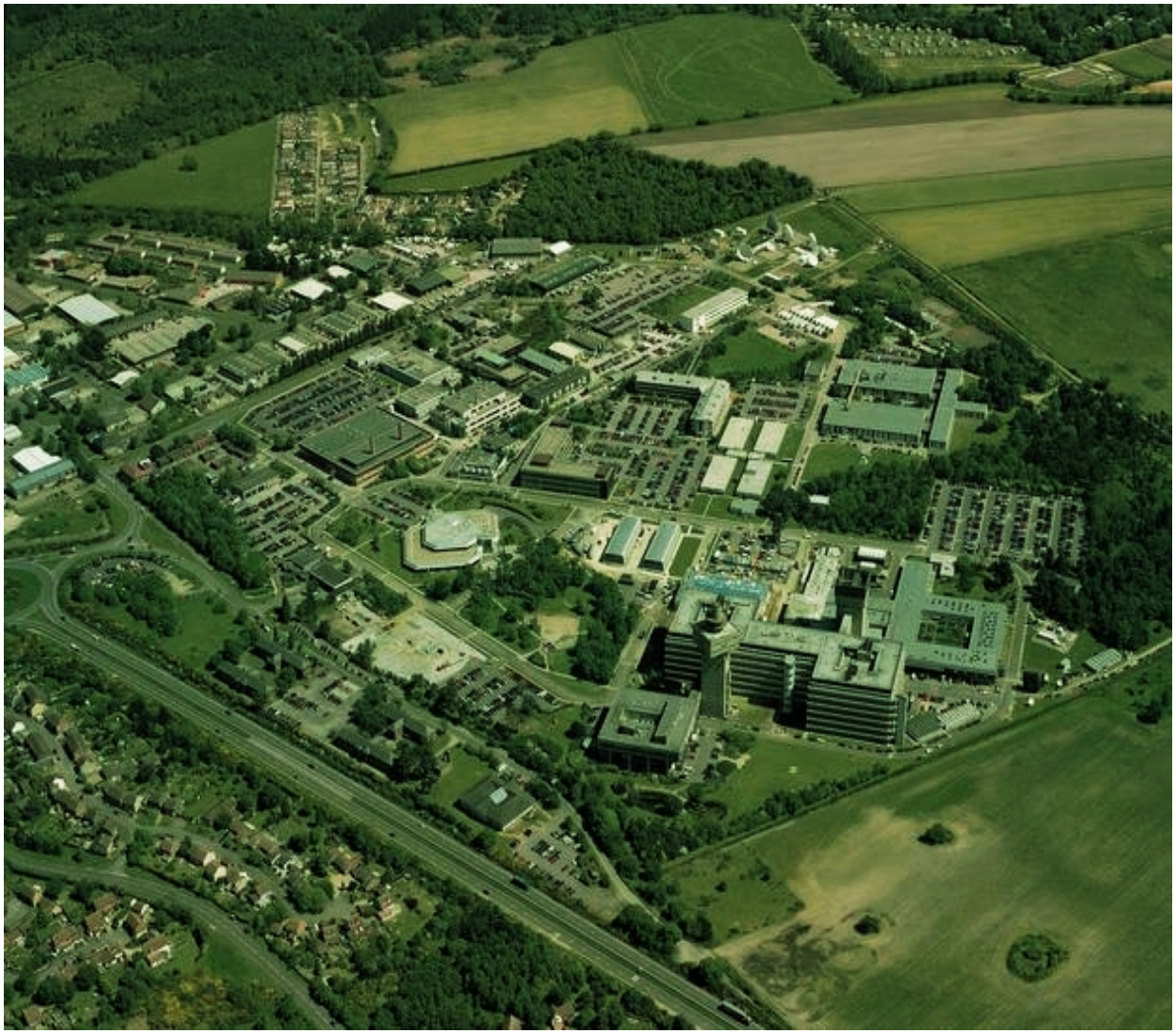
more.btexact.com/people/briggsk2



University of York MSc talk 2005 Feb 28 1415

typeset 2005 February 25 9:52 in pdfLaTeX on a linux system

Adastral Park, Martlesham, Suffolk



- Cambridge-Ipswich high-tech corridor
- 2000 technologists
- 15 companies
- UCL, Univ of Essex

BT Research centres

- Broadband centre
- Foresight centre
- IT futures research centre
- Intelligent systems centre
- Mobility centre
- Networks centre
- Pervasive ICT centre
- Security centre
- Asian Research centre
- Disruptive lab at MIT

MSc projects supervised by Keith Briggs

- Bayesian text processing
- exponential random graphs
- Lévy distribution in network traffic modelling
- analysis of internet autonomous systems graphs
- formal power series algorithms for combinatorics
- symplectic integrators

Bayesian text processing

- This subject has a fascinating combination of statistics and linguistics. It has recently become prominent as a result of its very successful use in spam filtering. This need has prompted rapid developments in the theory and practice.
- Another use is automatic correction of scanned documents converted to text by optical character recognition. This is especially challenging when the document contains mixed languages.
- Hence, I propose this MSc project:
  - *survey the theory, with an emphasis on what statistics can be done rigorously.*
  - *write a "multilingual spelling checker" that would operate without dictionaries. I imagine a program that highlights each word with a colour indicating the guessed language, with a special colour indicating a spelling error. In the latter case, the program would offer some suggested corrections.*

Exponential random graphs

- this is a class of random graphs proposed as models for social networks and now being applied to communication networks
- they are constructed with a very close analogy to statistical mechanical models in physics, so that one may talk of the temperature, free energy, entropy etc. of a graph!
- in this project we would write software to simulate this graph process, and study the problem of parameter estimation and fitting to observed graphs

Lévy distribution in network traffic modelling

- Lévy distributions are a class of random variables that have at least two attractive properties for network traffic modelling:
  - *aggregation property - the sum of two or more Lévy variables is also Lévy.*
  - *heavy tails, which are observed in network traffic.*
- this project would have three components:
  - *survey of the field, including Lévy processes.*
  - *writing of a program to fit parameters from the family of Lévy distributions (there are 4 parameters). Testing of the program on simulated data.*
  - *collection of internet traffic data and fitting of Lévy models. Discussion of quality of fitting.*

Analysis of internet autonomous systems graphs

- The millions of hosts forming the internet can be grouped into about 15,000 AS nodes. Daily data concerning these graphs is available. It is of interest to derive as much statistical information as possible from these graphs. Since they are large, this is challenging.
- The project would attempt:
  - *to refine and check what has already been done;*
  - *to compute other graph-theoretical descriptive statistics;*
  - *to critically review the value of these statistics.*
- If possible, a good conclusion would be to define random graph models which reproduce these statistics.

Formal power series algorithms for combinatorics

- *Suppose a_k is the number of object of size k in some collection, e.g. graphs with k nodes. In fact, let's take a narrower example - rooted trees, which come up often in computer science. The sequence a_k starts $1, 1, 2, 4, 9, 20, 48, 115, 286, 719, 1842, \ldots$ but how do we efficiently compute a_k for large k? It turns out that if*

  $A(x) = \sum_{k=0}^{\infty} a_k x^k,$

  then

  $A(x) = x \exp(\sum_{k=0}^{\infty} A(x^k)/k).$

- *This is a functional equation (FE), which implicitly defines A. We can solve it in various iterative ways, but a really good way would be a Newton method, which means that if at some stage we have computed n terms, the next iteration gives us 2*n terms. Such methods really do exist, but typically there is an overhead compared to simpler ways, which means that Newton methods do not always win in practice. The project would be to code as efficiently as possible various algorithms for the same FE, and compare the timings and memory requirements. Ultimately, we would like a library of algorithms which can easily be applied to FEs of this general type. Because the coefficients a_k can get very large, this really needs to be done in C, for which the best large-integer library (gmp) is available, and in which memory allocation can be controlled at a low level, which I have found to be necessary. Note: this is not numerical analysis in the traditional sense, because there is no floating point arithmetic.*

Symplectic integrators

- In recent years the theory of symplectic integrators for Hamiltonian ODE systems has become highly developed.
- However, there are not yet many practical codes available that are easy to use for non-experts. The project would take as starting point the Fortran package gnicodes.tar from and develop from it a new package in C (possibly with python etc. interfaces). This is an exercise in mathematical software design. Even the translation from f77 to C is not so easy, as dynamic memory allocation issues have to be sorted out.
- The final package should be in a standard format with auto-configuration, self-testing, and documentation (perhaps using doxygen). It would be good if the self-test program ran the examples from the book referred to above, and reproduced some of the work-accuracy diagrams therein.
- A further suggestion is to produce a version which could be submitted for inclusion in GSL

MSc projects supervised by Maziar Nekovee

- Internet on the road: Modelling and simulation studies of vehicular ad-hoc networks

- Vehicular ad-hoc networks are formed by vehicles equipped with short-range radio transmission capabilities. These networks are currently being intensively studied as means for extending the range of connectivity and internet coverage to cars on the road. The applications of VANET range from road traffic monitoring and control to music downloading and Internet browsing on the move.

  The aim of this project is to study, via mathematical modelling and simulations, the structure and dynamics of these networks. In particular, we aim to:

  - *build a realistic microscopic model of vehicular mobility in urban traffic settings, by using microscopic traffic models, and extending an existing in-house simulator of vehicular highway traffic*
  - *investigate the connectivity and other structural properties of the time-dependent of ad-hoc networks that emerge when these moving vehicles are equipped with short-range communication devices*
  - *simulate routing of packets and cooperative downloading in vehicular networks in urban traffic settings*

MSc projects supervised by Sverrir Olafsson