‘The First Day of Summer’: Parsing Temporal Expressions with Distributed Semantics

Ben Blamey, Tom Crick, Giles Oatley
Cardiff Metropolitan University

beblamey@cardiffmet.ac.uk
http://benblamey.name
@benblamey

Outline

- What are temporal expressions?
- How do we represent temporal expressions?
  - Existing Approach (Discrete Interval)
  - Issues
  - Proposed Approach – *Distributed Definition*
- Mining the Definitions
  - Collect Data
  - Gaussian Mixture Models & Expectation-Maximization
- Parsing – adaption of StanfordNLP / SUTime.
- Examples / Online Demonstration
- Application – when was my photo uploaded?

What are Temporal Expressions?

"Next Tuesday is my Halloween Party"

"Can't wait for Bonfire Night"

"Today is Chinese New Year"

"I'm moving house at the start of January"

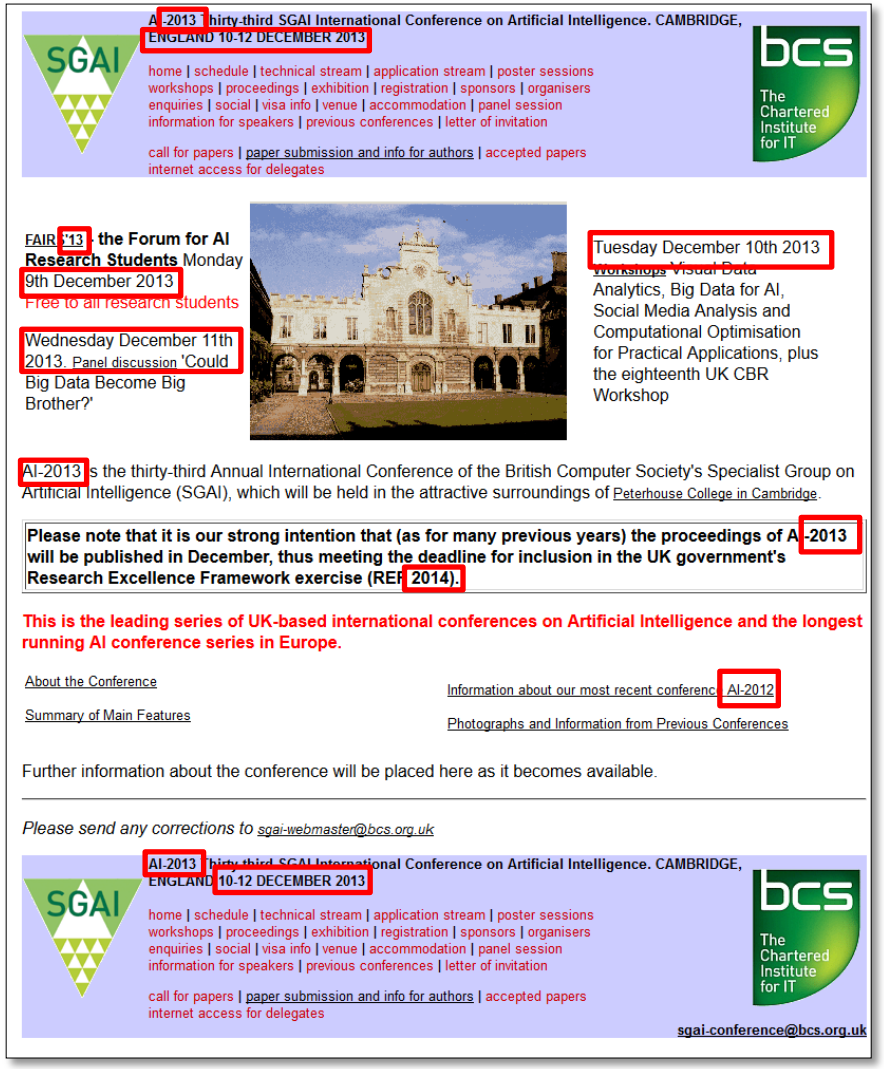
FRESHERS WEEK!!!

"Last Summer I really enjoyed the London Olympic Games"

3/4/2014

2013-08-09 12:45 UTC+5

"On my way to AI-2013"



AI-2013 Thirty-third SGAI International Conference on Artificial Intelligence. CAMBRIDGE, ENGLAND 10-12 DECEMBER 2013

home | schedule | technical stream | application stream | poster sessions workshops | proceedings | exhibition | registration | sponsors | organisers enquiries | social | visa info | venue | accommodation | panel session information for speakers | previous conferences | letter of invitation

call for papers | paper submission and info for authors | accepted papers internet access for delegates

FAIR'13 the Forum for AI Research Students Monday 9th December 2013 Free to all research students

Wednesday December 11th 2013. Panel discussion 'Could Big Data Become Big Brother?'

Tuesday December 10th 2013 Workshops Visual Data Analytics, Big Data for AI, Social Media Analysis and Computational Optimisation for Practical Applications, plus the eighteenth UK CBR Workshop

AI-2013 is the thirty-third Annual International Conference of the British Computer Society's Specialist Group on Artificial Intelligence (SGAI), which will be held in the attractive surroundings of Peterhouse College in Cambridge.

Please note that it is our strong intention that (as for many previous years) the proceedings of AI-2013 will be published in December, thus meeting the deadline for inclusion in the UK government's Research Excellence Framework exercise (REF 2014).

This is the leading series of UK-based international conferences on Artificial Intelligence and the longest running AI conference series in Europe.

About the Conference

Summary of Main Features

Information about our most recent conference AI-2012

Photographs and Information from Previous Conferences

Further information about the conference will be placed here as it becomes available.

Please send any corrections to sgai-webmaster@bcs.org.uk

AI-2013 Thirty-third SGAI International Conference on Artificial Intelligence. CAMBRIDGE, ENGLAND 10-12 DECEMBER 2013

home | schedule | technical stream | application stream | poster sessions workshops | proceedings | exhibition | registration | sponsors | organisers enquiries | social | visa info | venue | accommodation | panel session information for speakers | previous conferences | letter of invitation

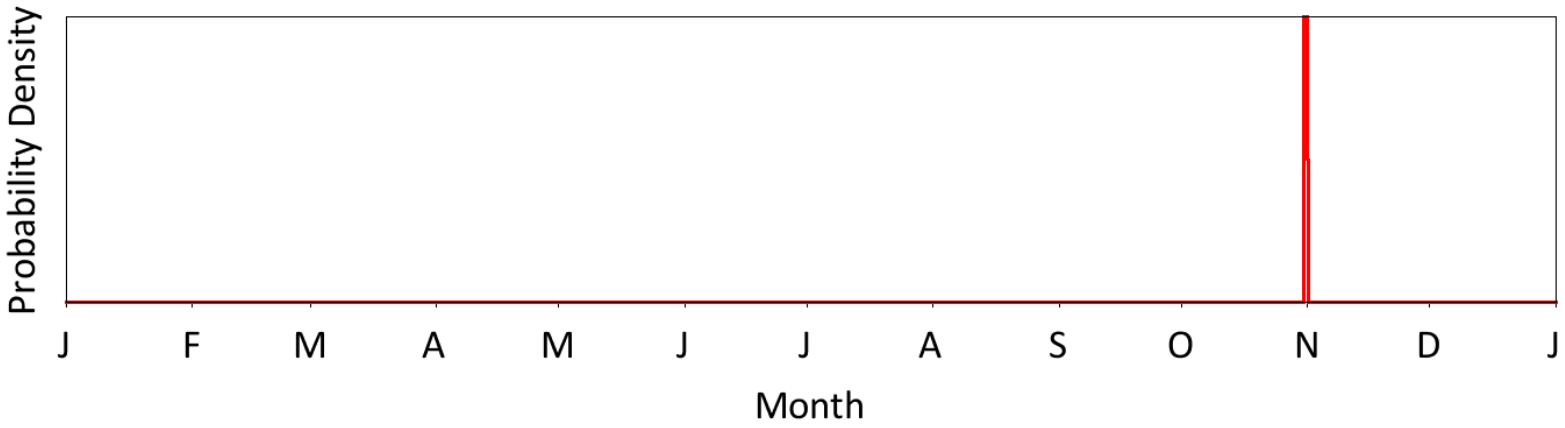
call for papers | paper submission and info for authors | accepted papers internet access for delegates

sgai-conference@bcs.org.uk

How do we Represent Temporal Expressions?

"Can't wait for my Halloween party"

Discrete Interval (Traditional Approach)

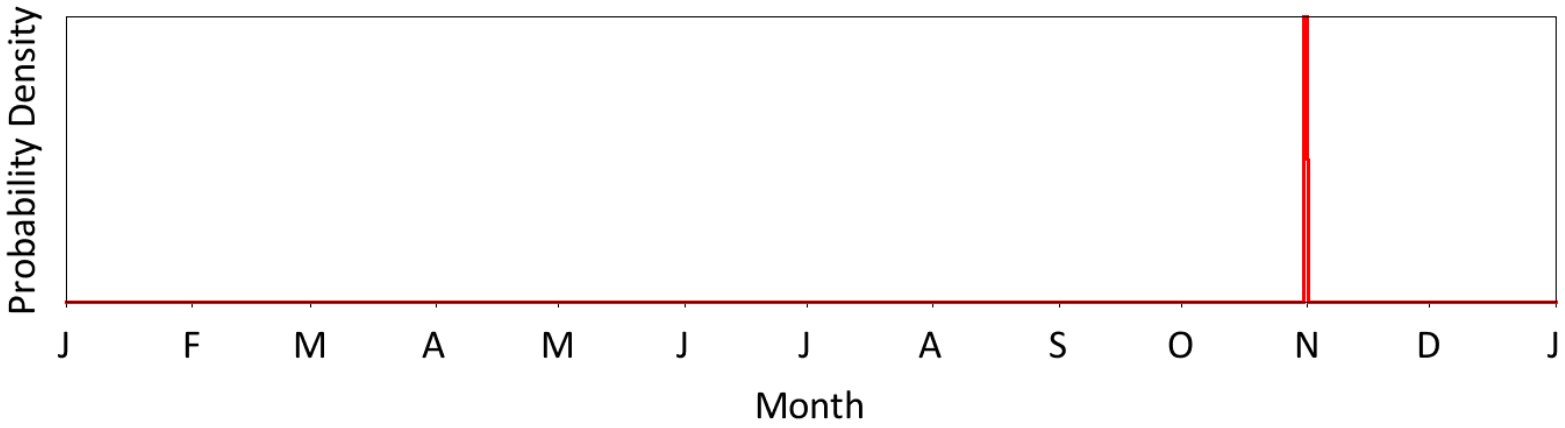



```
<TIMEX3 tid="t1" type="DATE"
        value="2013-10-31"
        range="(2013-10-31,2013-10-31,P1D)">
```

How do we Represent Temporal Expressions?

"Can't wait for my Halloween party"

Discrete Interval (Traditional Approach)




```
<TIMEX3 tid="t1" type="DATE"
        value="2013-10-31"
        range="(2013-10-31,2013-10-31,P1D)">
```

Was your party definitely Midnight->Midnight on 31st? No.

=> Here, this approach gives us a poor representation of the *real meaning*.

Key Idea

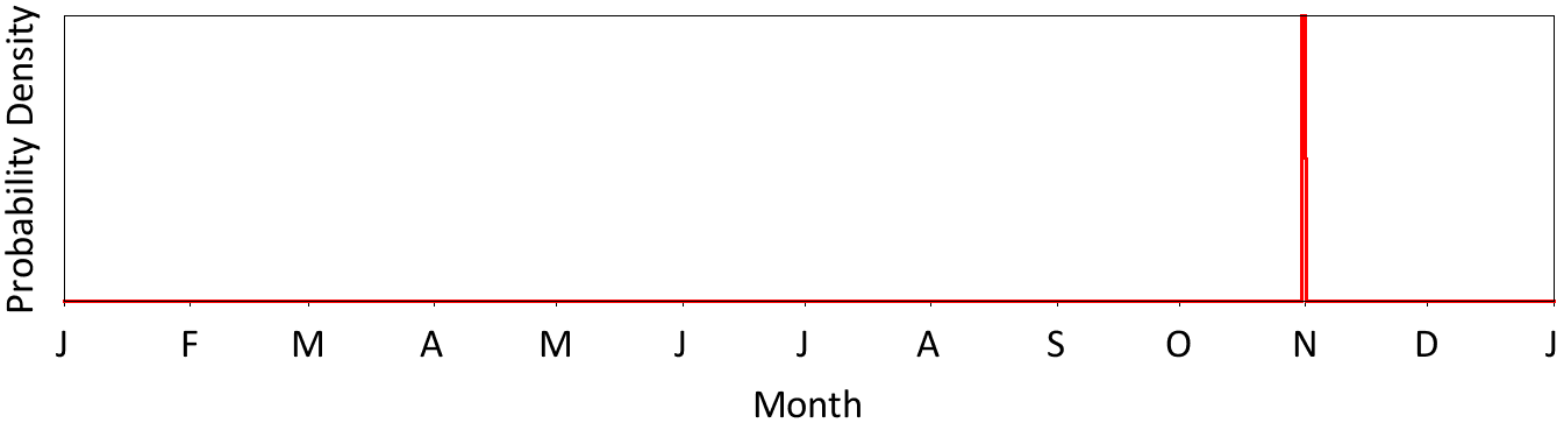
We can combine lots of definitions
to build a statistical distribution of these definitions.

...and use this the distribution
as a new definition.

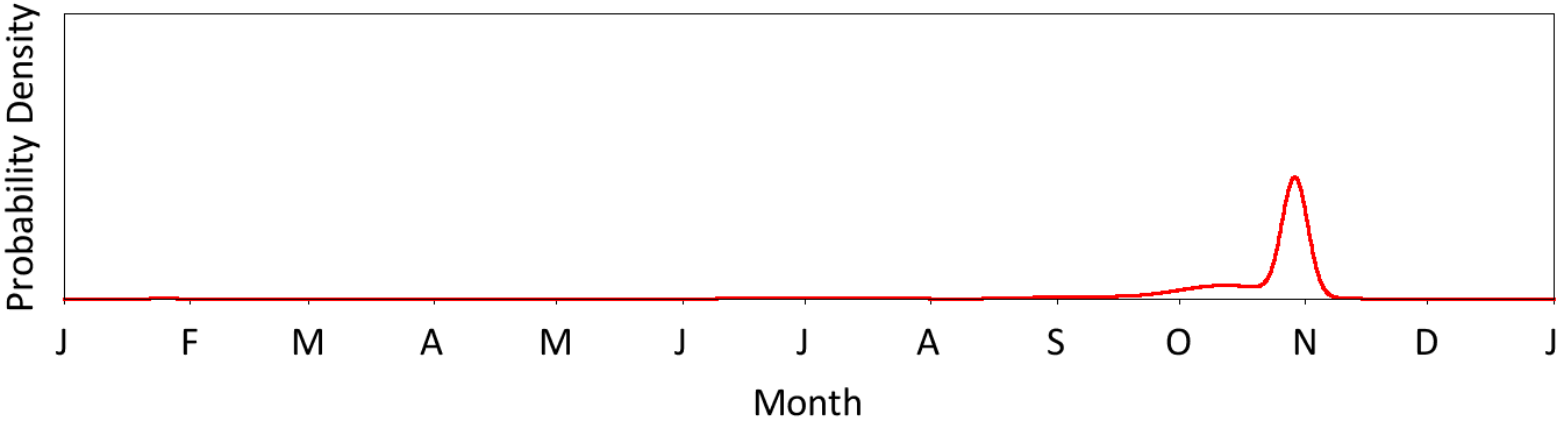
How do we Represent Temporal Expressions?

"Can't wait for my Halloween party" (31st October)

Discrete Interval:

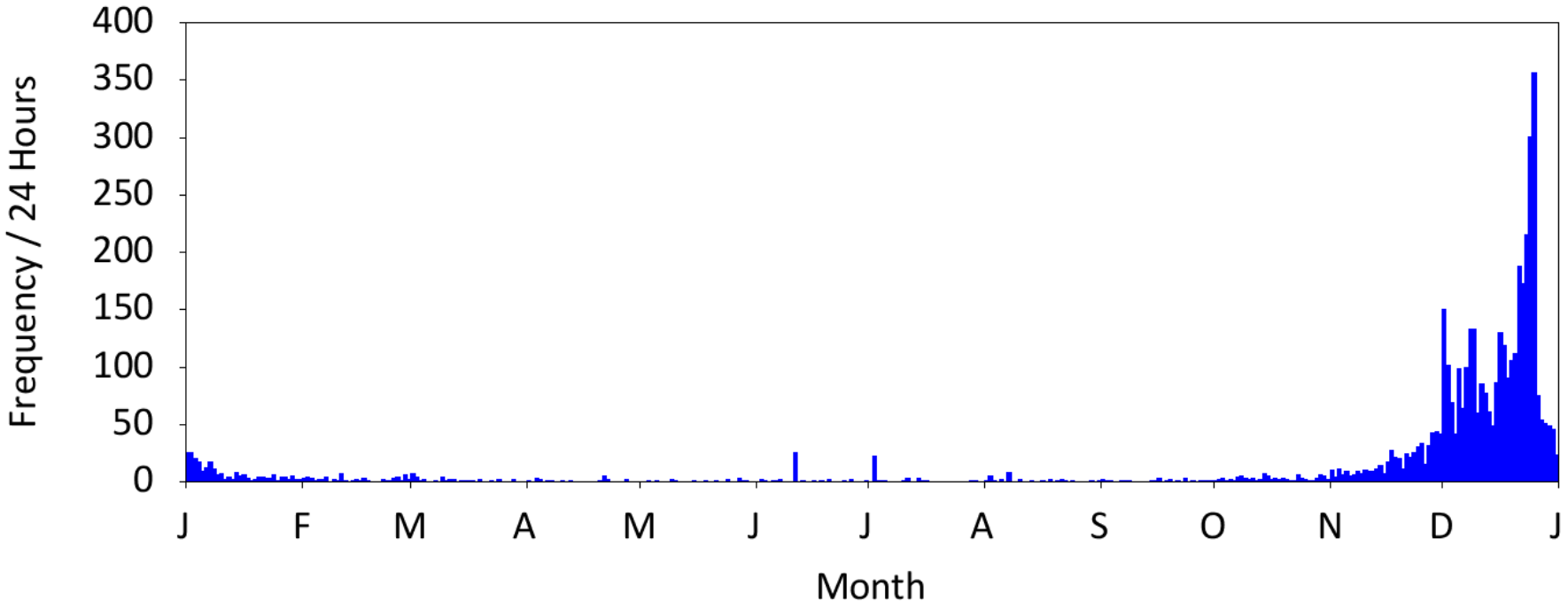


Distributed:



Data Collection – "Christmas"

1. Search Flickr for "Christmas" photos uploaded in 2012.
2. Extract the timestamp from the EXIF metadata.
3. Plot a frequency distribution from these results:

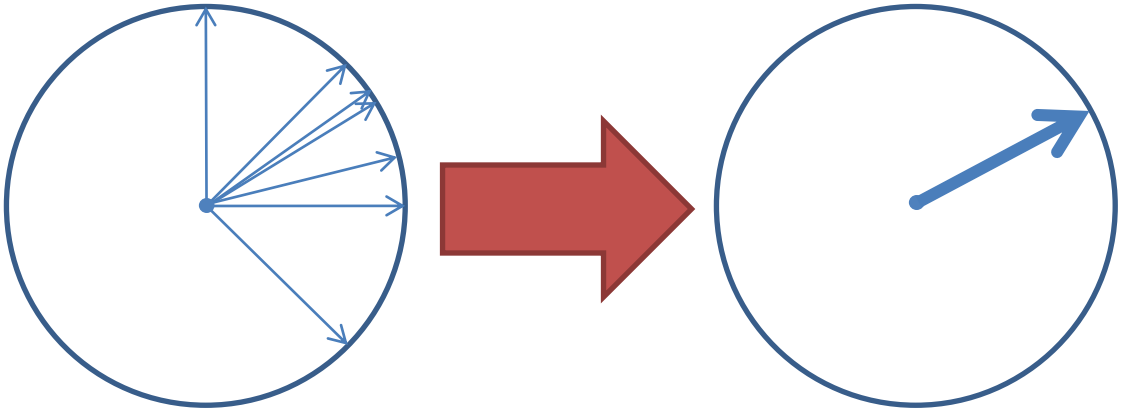


beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

Fitting – Mathematical Model

- Consider the year as a circular quantity.
- Modulo arithmetic in the usual way.
- *mean of circular quantities*

  (think: resolving forces in Newtonian Physics)



Fitting – Gaussian Mixture

- Gaussian Mixture Model just means the distribution is a weighted sum (or mixture) of more than one Normal Distribution.
- Include a Uniform Distribution to obviate distortion from “background noise”.

$$T_s(t) = a_0 k + \sum_{i=1}^{i=M} a_i N(\mu, \sigma^2)$$

where $\sum a_i = 1$, $a_i \geq 0$

and $\int_{Year} T_s(t) = 1$

t = The datetime, under some unit scheme, where the start of the year has value 0.

M = The number of normal distributions in the mixture.

$N(\mu, \sigma^2)$ = Normal Distribution for mean μ and variance σ^2.

a_i = The weight given to distribution i

$T_s(t)$ = The probability density function associated with the temporal expression.

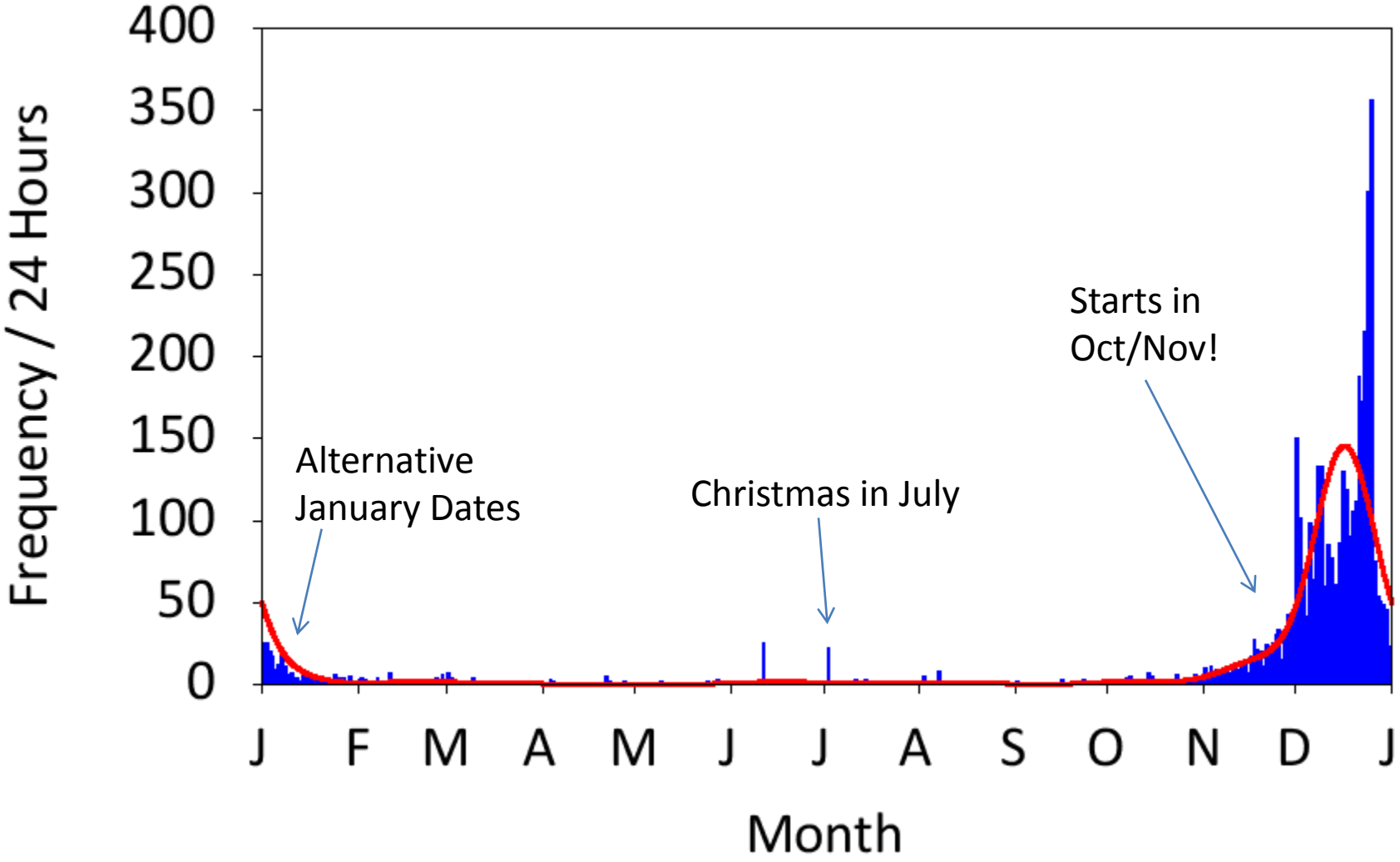
k = 'height' of uniform distribution (can easily be calculated based on our unit scheme)

Fitting – Expectation Maximization

```
Initialize the mixture model params: {{weight,mean,variance}}

DO {
    // EXPECTATION STEP (model -> data)
    foreach (Frequency in observations) {
        foreach (Model in mixture) {
            γ[][] <- weight * model(observation).
        }
    }

    // MAXIMIZATION STEP (data -> model)
    foreach (Model in mixture) {
        Estimate new model mean, variance & weight from
          the γ values for that model.
        // (Models that made little overall contribution
        // have their weight reduced accordingly.)
    }

    // CHECK FOR CONVERGENCE
    Compute new loglikelihood.
    continue = (#iterations < limit) && (Δloglikelihood > threshold)
} WHILE (continue)
```

http://code.google.com/p/accord/

Example – "Christmas"



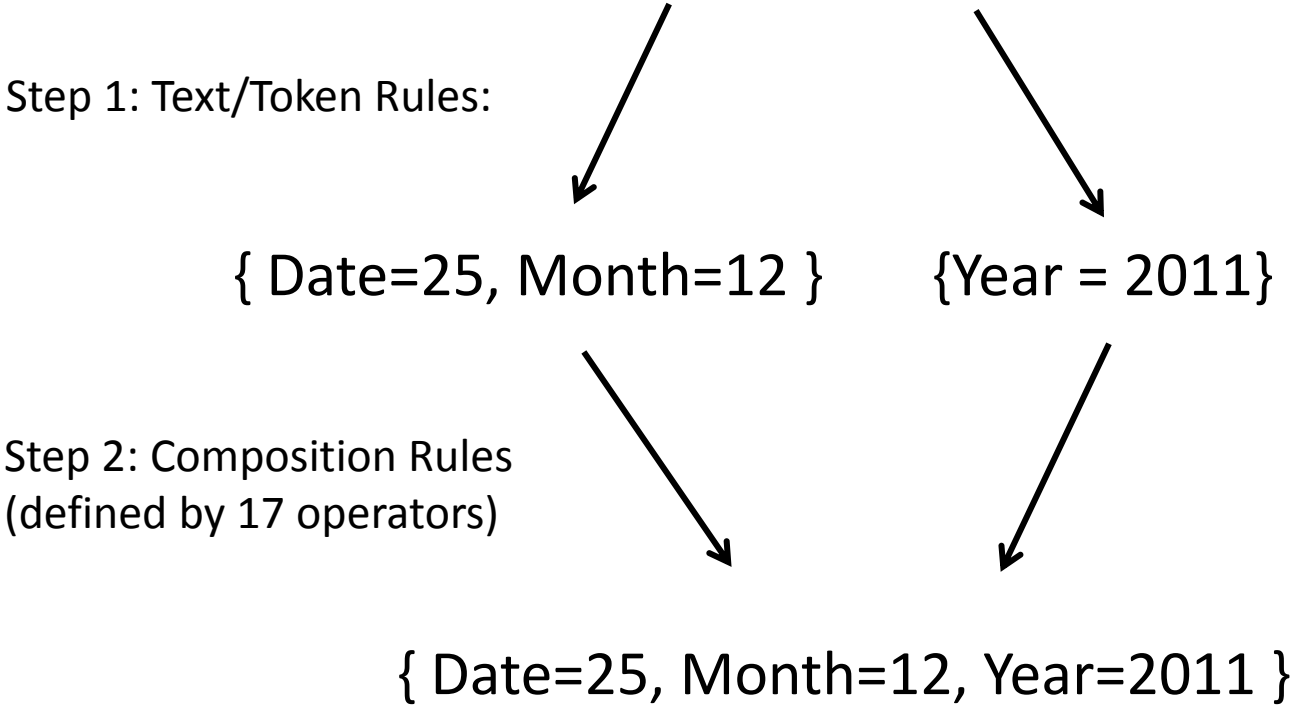
beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

SUTime - Overview

- State-of-the-art Tempex Annotator developed at Stanford University. Part of the StanfordNLP toolkit.
  - http://nlp.stanford.edu/software/corenlp.shtml
- Mixture of Java code and grammar files defined in the JSON-like TokensRegex format.

SUTime - Parsing

"I took this photo in Christmas 2011."



Read more here: http://nlp.stanford.edu/software/tokensregex.shtml

Adapting the Grammar - Rules

- Define new token-level rules.
  - Many temporal expressions like “fresher’s week” simply don’t have a discrete-interval representation, existing definitions are augmented.

```
{ (/fresher'?s/ /week/ ) => PdfTime(FRESHERSWEEK_DIST) }
{ (/bonfire/ /night/ ) => IsoDate(NIL, 12, 31, BONFIRENIGHT_DIST) }
{ (/st.?|saint/? /valentine/ $POSS? /day/ ) => IsoDate(NIL, 2, 14, VALENTINESDAY_DIST) }
```

- The parameters for the mixture models are exported:

```
FRESHERSWEEK_DIST = SumTimeExpression(
    AnnualNormalDistribution(0.620141003011485,402030378.658,640202.171704994),
    AnnualNormalDistribution(0.234572698142028,401109640.787,171650.789160729),
    AnnualNormalDistribution(0.0505045144189651,405853428.672,663596.589699629),
    AnnualNormalDistribution(0.047940522646717,382816921.807,216.8672694097),
    AnnualNormalDistribution(0.0214286670877024,380634533.557,18836.7068154974),
    AnnualUniformDistribution(0.0139941585850157),
    AnnualNormalDistribution(0.00701558167005665,409170117.67,1830918.09260277),
    AnnualNormalDistribution(0.00266863095904928,396513718.609,200066.441399321),
    AnnualNormalDistribution(0.00133392369614968,392324582.8,701969.432955515)
)
```

Representation of P.D.F

- Adapt the Temporal classes to give them additional probability-density definition.
- Create classes to represent the composition of the functions:

AnnualUniformDistribution, AnnualNormalDistribution
SumTimeExpression, IntersectTimeExpression

```
public interface ITimeDensityFunction {
    public double GetDensity(DateTime time);
    public String GetGNUPlot(
      String millTimeSecondsExpr);
}
```

Adapting the Grammar

- Implement calculation for `INTERSECT_OP` operation
  - The p.d.f. dual of intersection is multiplication of the functions.
  - Fallback to discrete interval representation as necessary (i.e. uniform distribution over the interval).

Online Demo

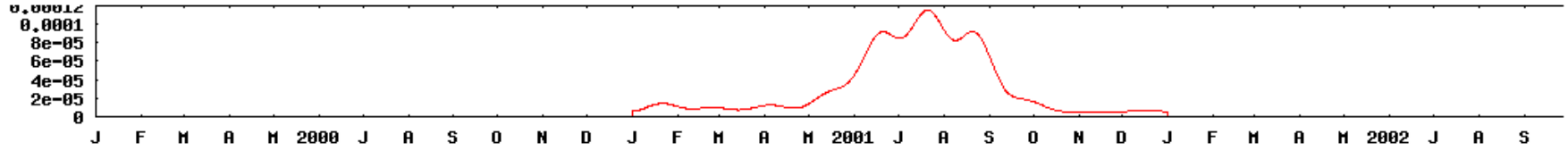
The First Day of Summer: Parsing Temporal Expressions with Distributed Semantics

Summer 2001

Submit

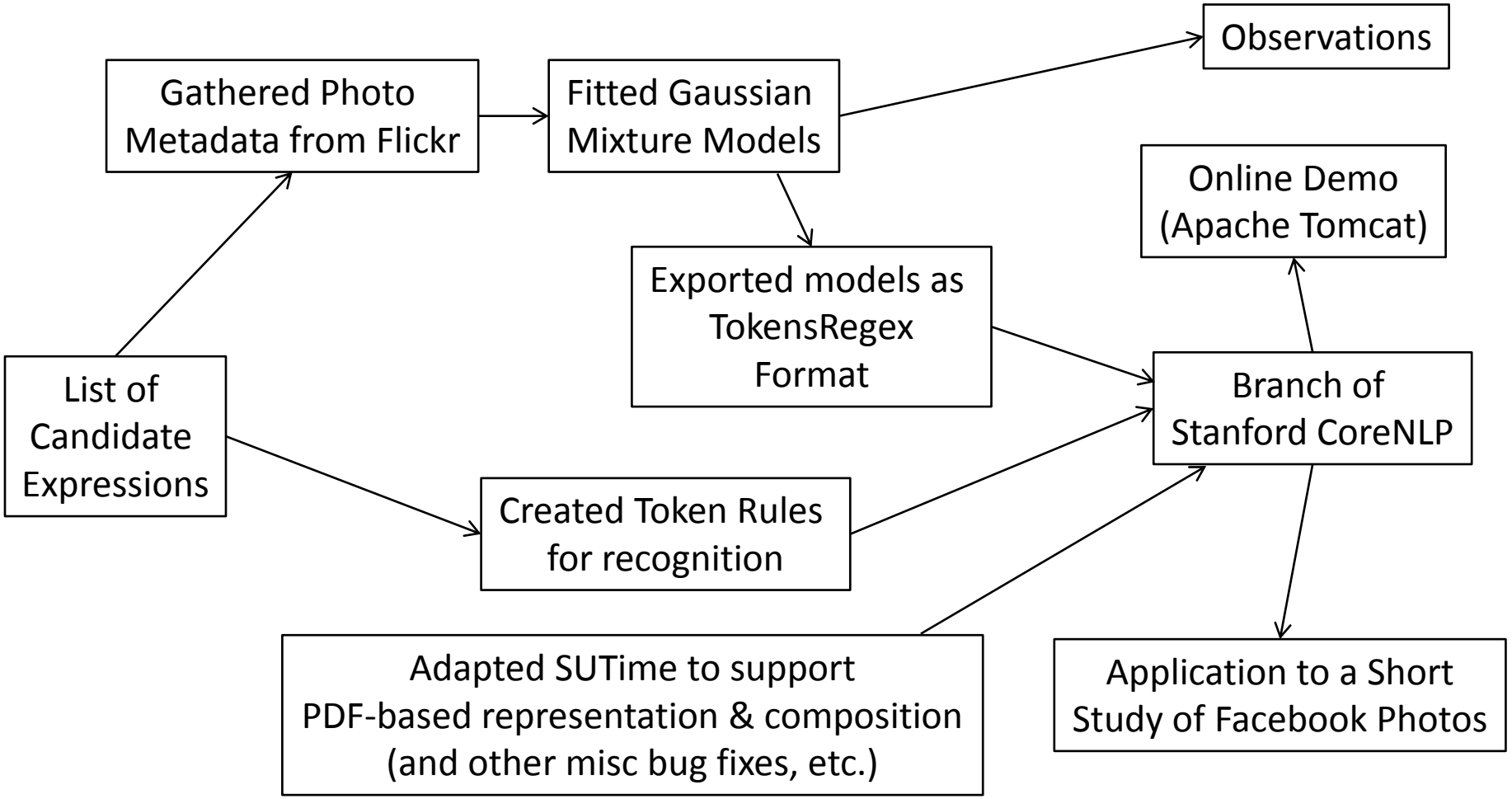
Summer 2001

<TIMEX3 X-GNUPlot-Function="(((0.293868748378889/(sqrt(1116628.25886232*2*pi))*exp (- dsyear(x , 3.96083115724E8)**2 / (2 * 1116

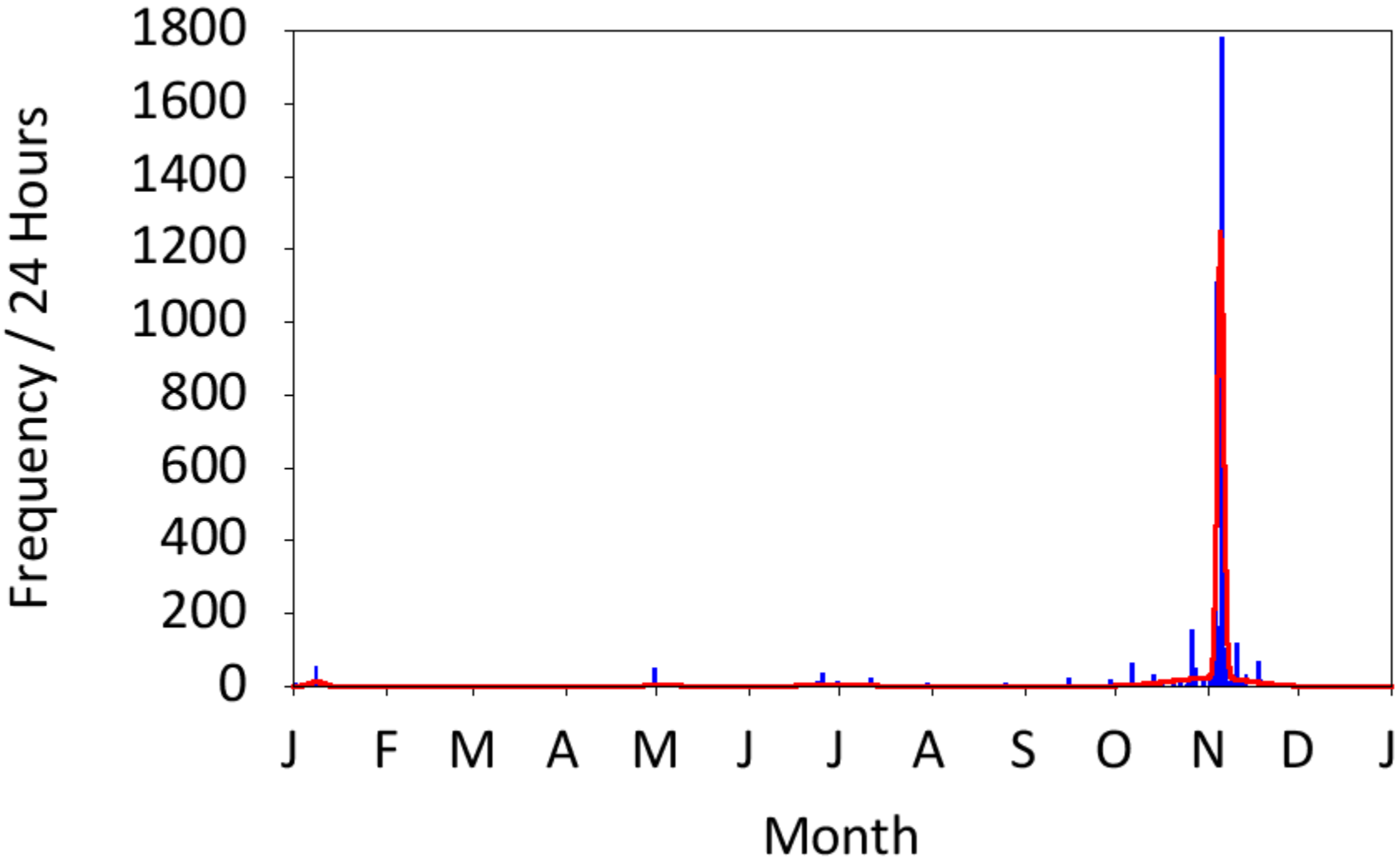


http://benblamey.name/

Overview of Process

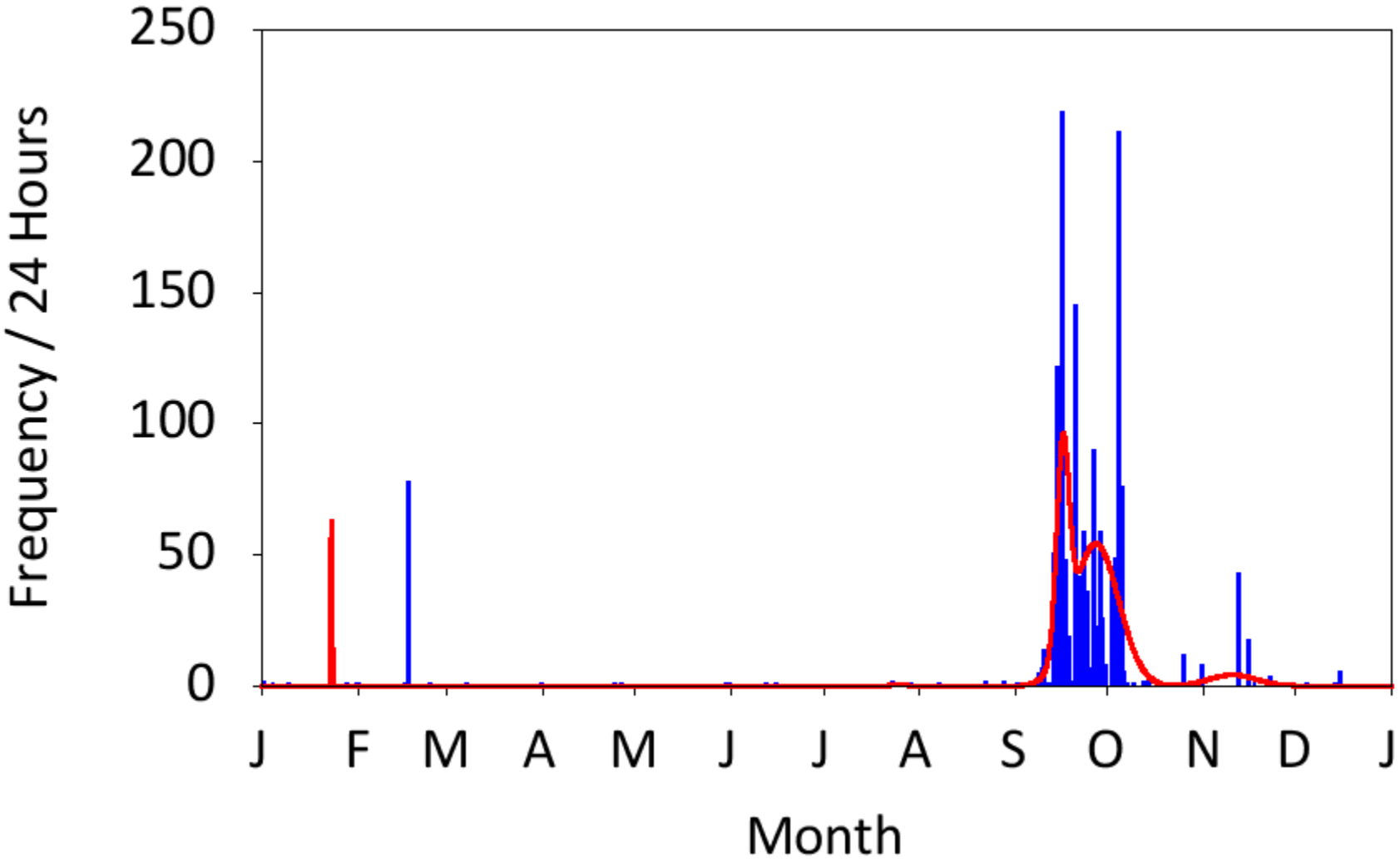


“Bonfire Night”



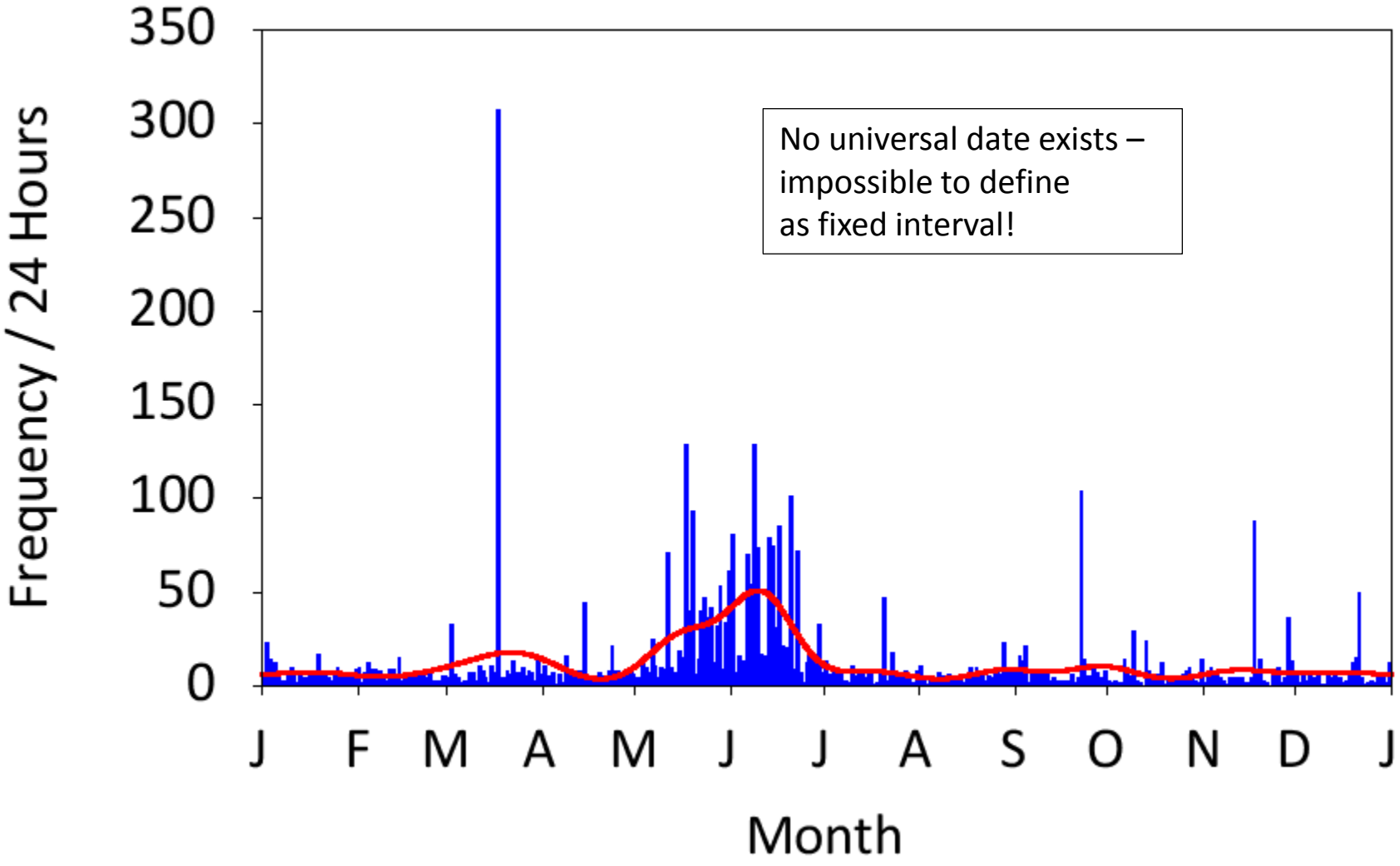
beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

“Fresher’s Week”



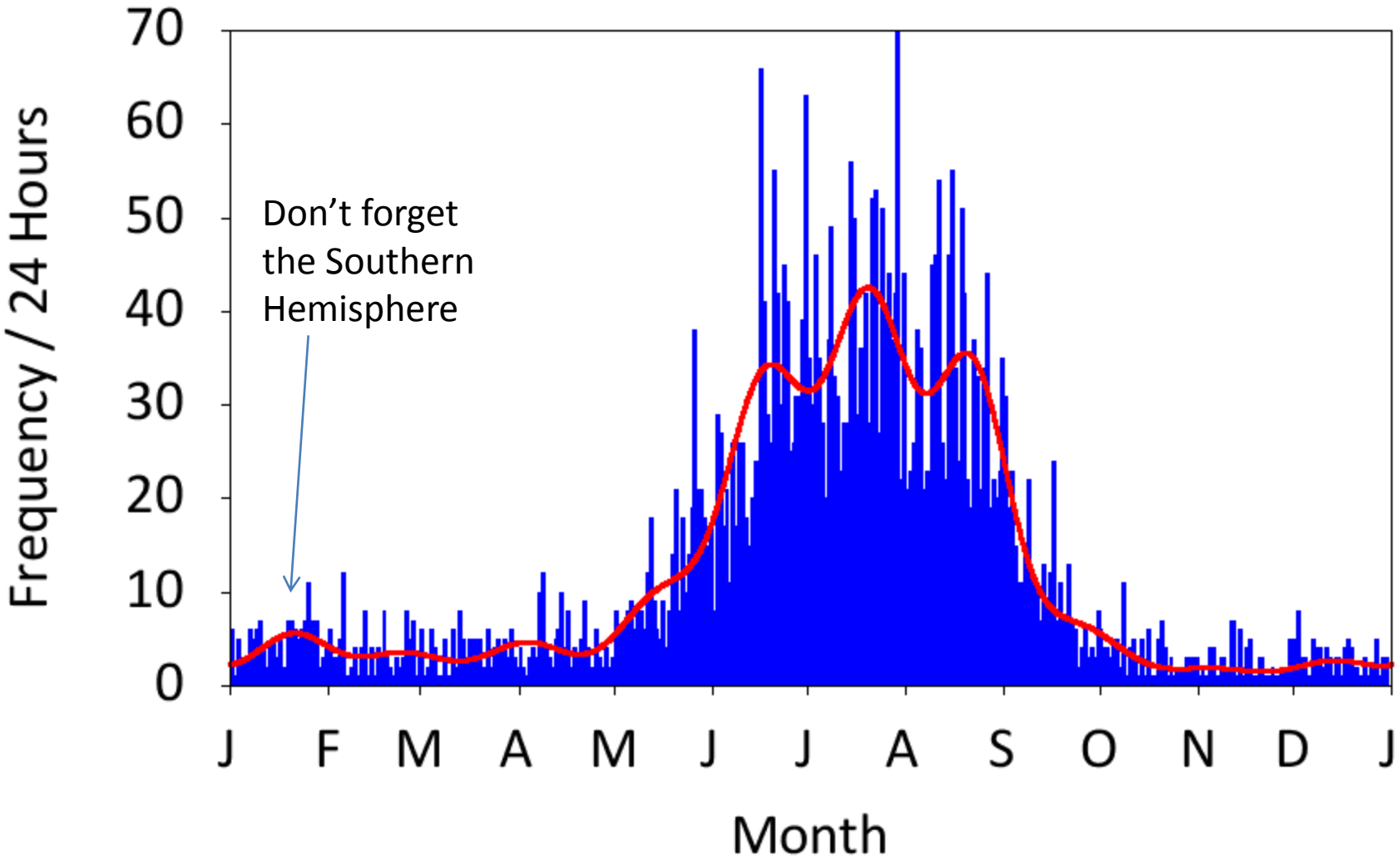
beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

“Last Day of School”



beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

“Summer”

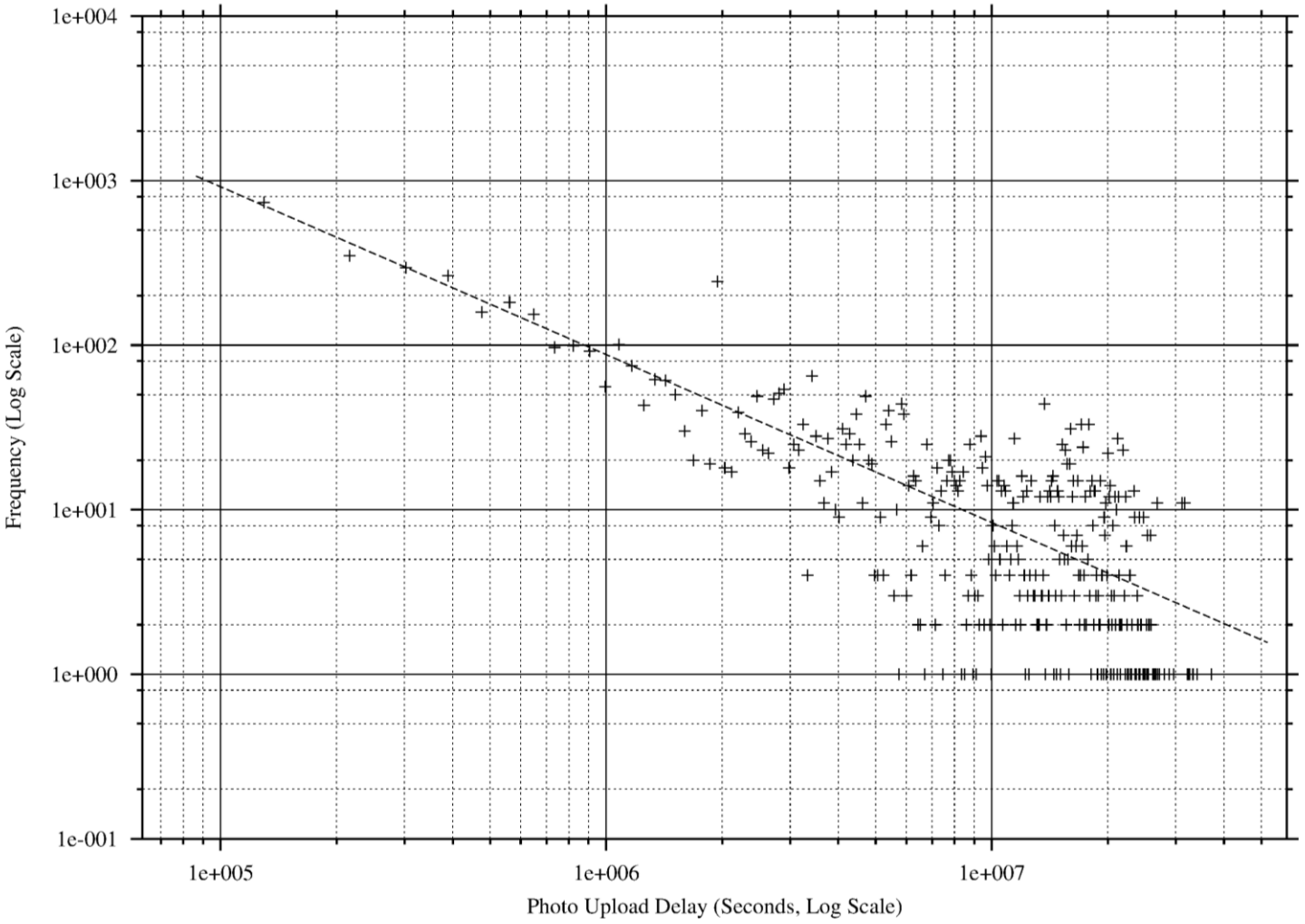


beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

Application: Facebook Photo Creation Time

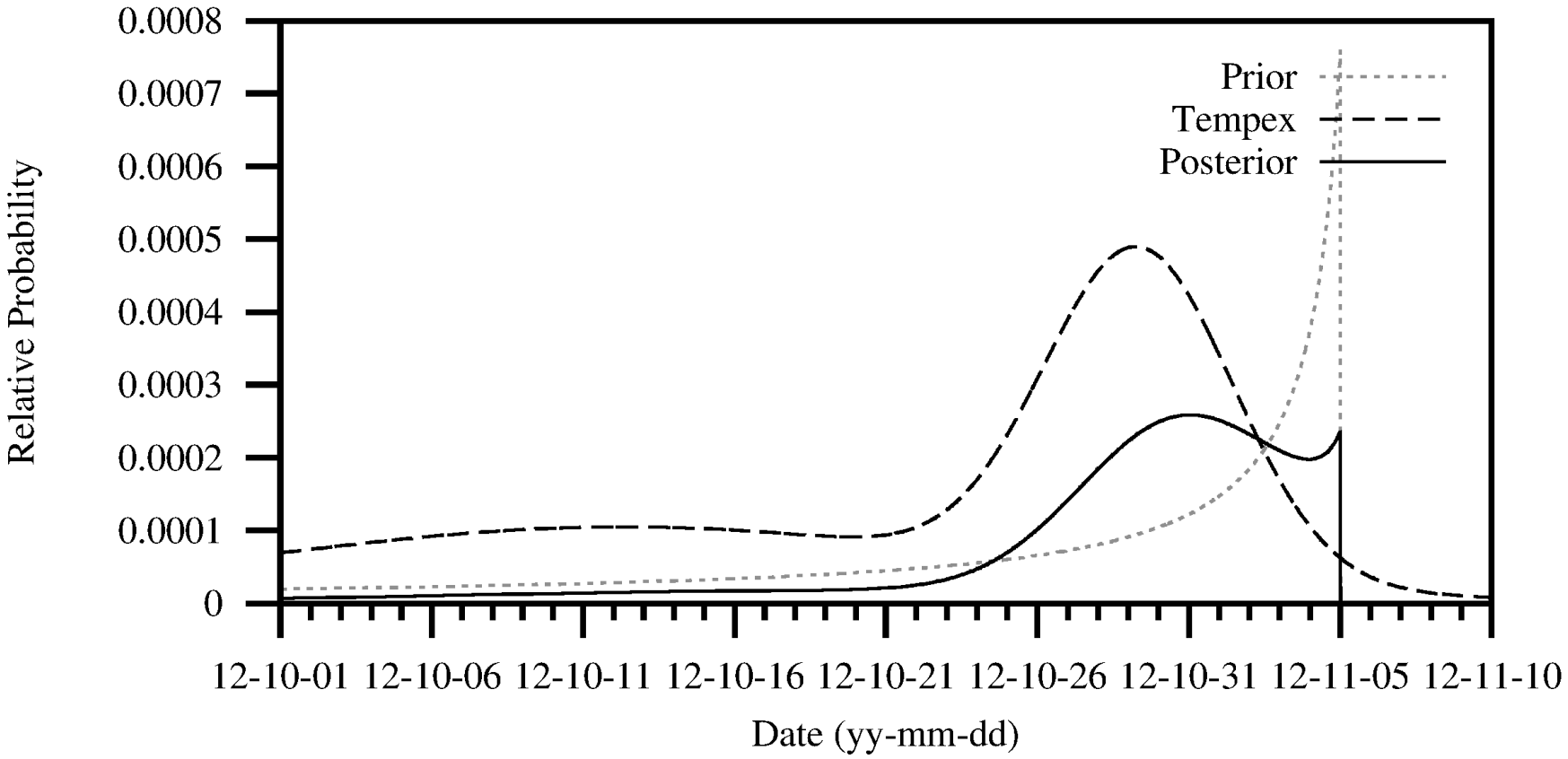
- Facebook Photos don't have EXIF.
  - We want to estimate when the photos were taken.
- We can build a prior distribution of the "photo upload delay" using the Flickr photos.
- The title of the photo might contain a temporal expression.
- We can combine our prior with the PDF yielded by parsing the temporal expressions.

Photo “Upload Delay”



beblamey@cardiffmet.ac.uk http://benblamey.name @benblamey

“Halloween”


0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0
Relative Probability
Prior
Tempex
Posterior
12-10-01
12-10-06
12-10-11
12-10-16
12-10-21
12-10-26
12-10-31
12-11-05
12-11-10
Date (yy-mm-dd)

Conclusions

Advantages of Distributed Definition:

- Wider scope of temporal expressions
  - Includes expressions where there isn't an "official" date e.g. "last day of school"
- It captures greater cultural richness and ambiguity – arguably a more 'accurate' definition.
  - "Bonfire night", "Christmas" – rich mixture of meanings.
- Motivates probability-based approaches to temporal similarity, and further calculations.
  - See the paper for a short study on computing photo upload time.

- Demonstrated techniques for data collection and fitting.

Questions?

‘The First Day of Summer’: Parsing Temporal Expressions with Distributed Semantics

Ben Blamey
Cardiff Metropolitan University

beblamey@cardiffmet.ac.uk
http://benblamey.name
(Source Code, Data, More Graphs, Online Demo)

If you want help using the code – get in touch!

Twitter: @benblamey