Bursty conversations on Twitter

A non-markovian stochastic process to model online conversations and analyse the structure of social networks

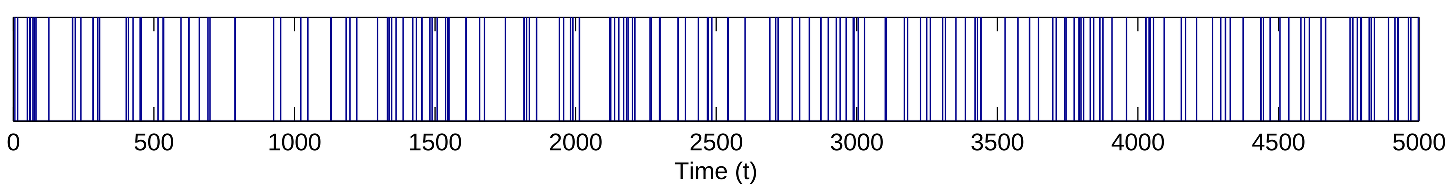
Ewan Colman and Danica Vukadinović Greetham
University of Reading
Ewan.Colman@Reading.ac.uk + D.V.Greetham@Reading.ac.uk

University of Reading
Centre for the Mathematics of Human Behaviour

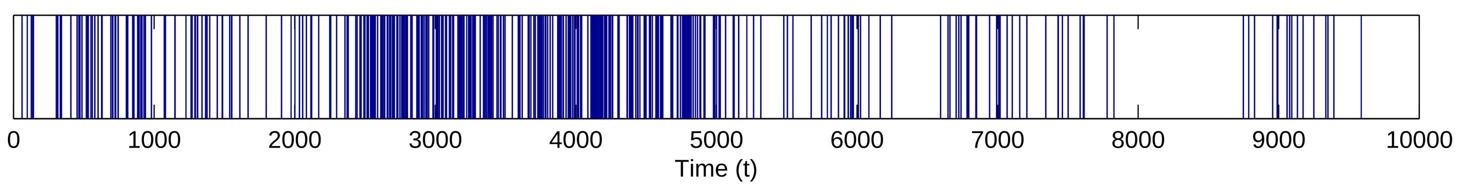
On this poster:

- A discrete-time stochastic process is described which produces a power-law inter-event time distribution with a tunable exponent. When these events are Tweets or emails the process becomes a model for human communication.
- Some empirical data-sets are shown to have inter-event time distributions which are consistent with the results of our model.
- A method is proposed (inspired by the model) for assessing the true weight of the relationship between any two individuals.
- When the network is weighted using this new method, it is shown to be more steady over time (and therefore a better predictor) than networks weighted by the total number of messages sent.

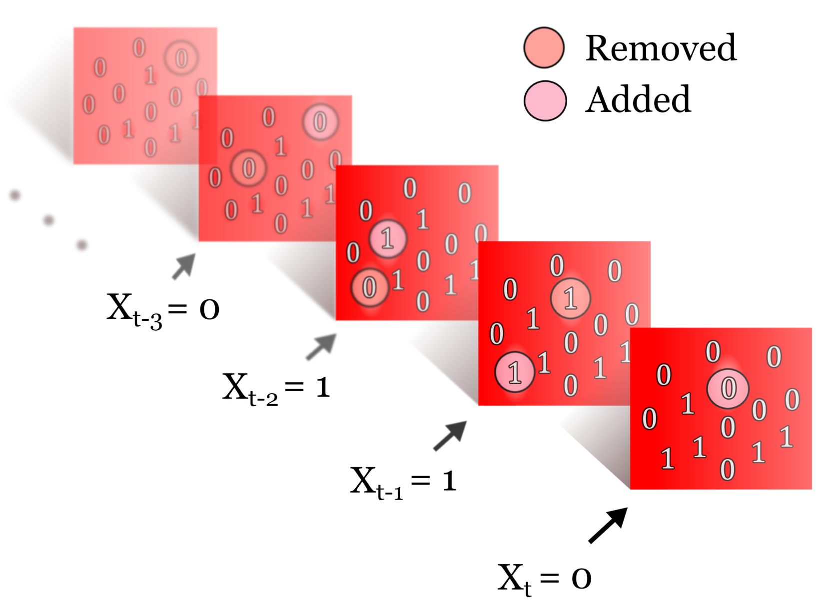
Mathematically, online communication between two people can be viewed as a discrete sequence of events; in each second of the day a message either will, or will not, be sent from person A to person B. If we assume that the probability that A will send a message to B is the same at every moment in time, say p, then the sequence of events will be described by a Bernoulli process. The probability that the *inter-event time* (or waiting time) between two events will be exactly τ is just $(1-p)^{\tau}$ and the inter-event times are **exponentially** distributed. The sequence would look something like this:



In reality this is not how people behave. Many messages are sent in response to messages received; there is an element of memory and feedback in the process. The burst-like quality of online conversations is characterised by the *power-law* inter event time distribution which appears as follows:



Here we can see long periods of inactivity punctuated by periods of high frequency communication. The model described below reproduces this behaviour. This new model of human dynamics has implications for the way social network data should be interpreted.



Generating the event sequence

We let X_t be a sequence of binary variables and declare that an 'event' is an occurrence of $X_t = 1$, whereas $X_t = 0$ is a moment of inactivity. k_t is the number of events stored in a memory bank (the red square), and M is the size of the memory. At each iteration a randomly selected (or alternatively the oldest) entry in the memory is overwritten by X_t, which itself is chosen as follows:

- With probability $f(k_t)$, $X_t = 1$. With probability $1 - f(k_t)$, $X_t = 0$.

Linear proportionality

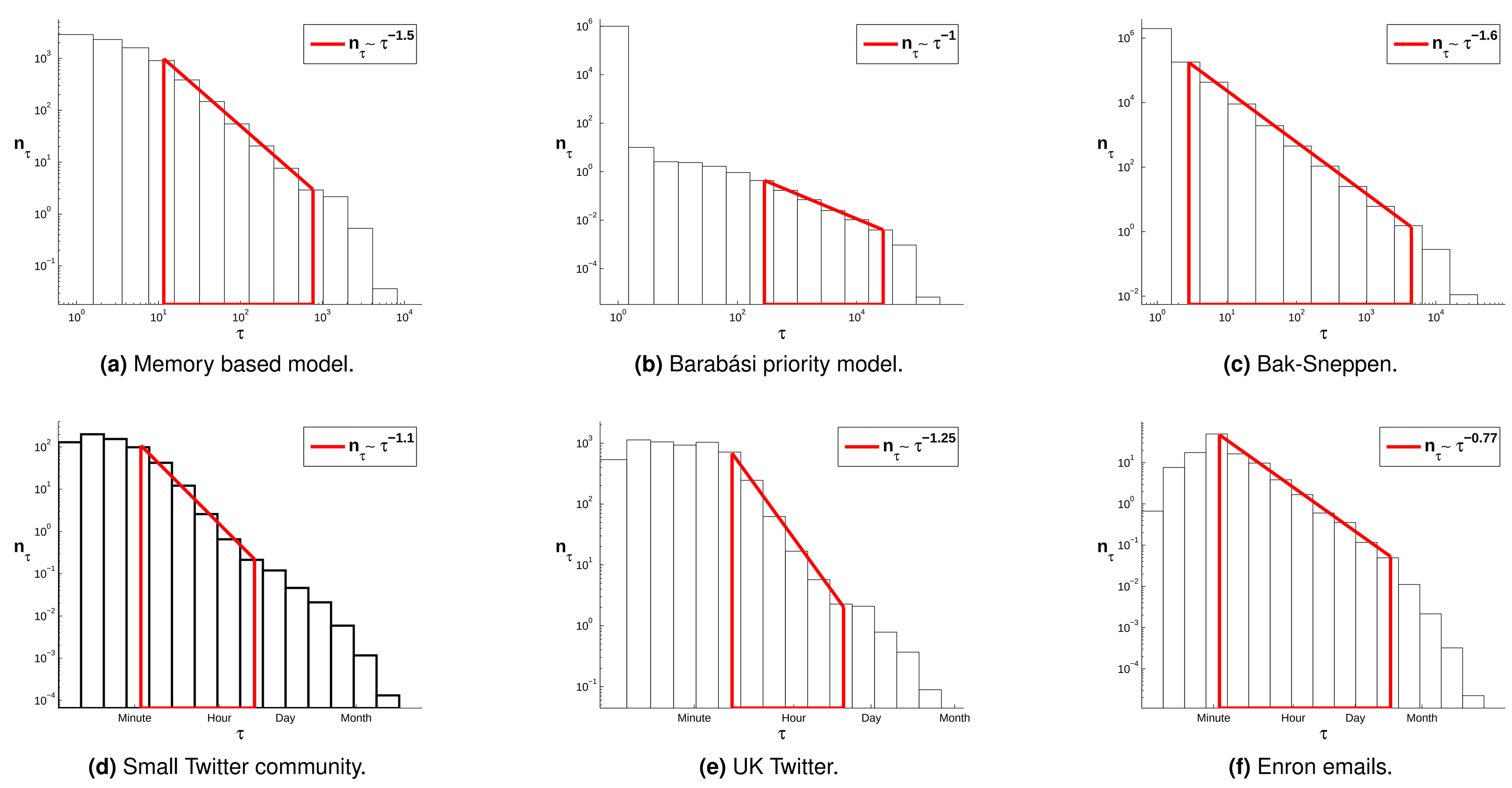
If we choose the following linear form for f,

$$f(k_t) = \begin{cases} f_0 & \text{if} \quad k_t = 0 \\ \dfrac{k_t + \delta}{M + \delta + 1} & \text{if} \quad k_t \geq 1 \end{cases}$$

then we find that the inter-event time distribution follows a power-law for $\tau < M$ (analytical solution omitted).

The probability Π_τ, that the length of a randomly selected interval between two events is exactly τ, is

$$\Pi_\tau \sim \tau^{-(2+\delta)}$$



(a) Memory based model. **(b)** Barabási priority model. **(c)** Bak-Sneppen.

(d) Small Twitter community. **(e)** UK Twitter. **(f)** Enron emails.

Figure 1: (1a) shows the results from a simulation of our model (with $M = 10^3$, $\delta = -0.5$, $f_0 = 10^{-3}$) along with simulations of two other computational models which are known to give power-law inter event time distributions. n_τ is the number of intervals of length τ. The bottom row shows the aggregated and log-binned inter-event times of a community of (1d) 28 Twitter users (1e) all Twitter mentions in the UK over a one month period, (1f) the Enron email data-set. Three separate regimes can be observed; firstly on the very short time-scale, where messages are sent in quick succession; second, there is a power-law distribution of waiting times; thirdly, on very large time-scales the inter-event times become exponentially distributed as we would expect for a Bernoulli process. In the Twitter data the pow-law occurs between around 1 minute and 8 hours, this is consistent with the memory based process when we set M, the length of the memory of the conversation, to be 8 hours.

Ironing out the bursts

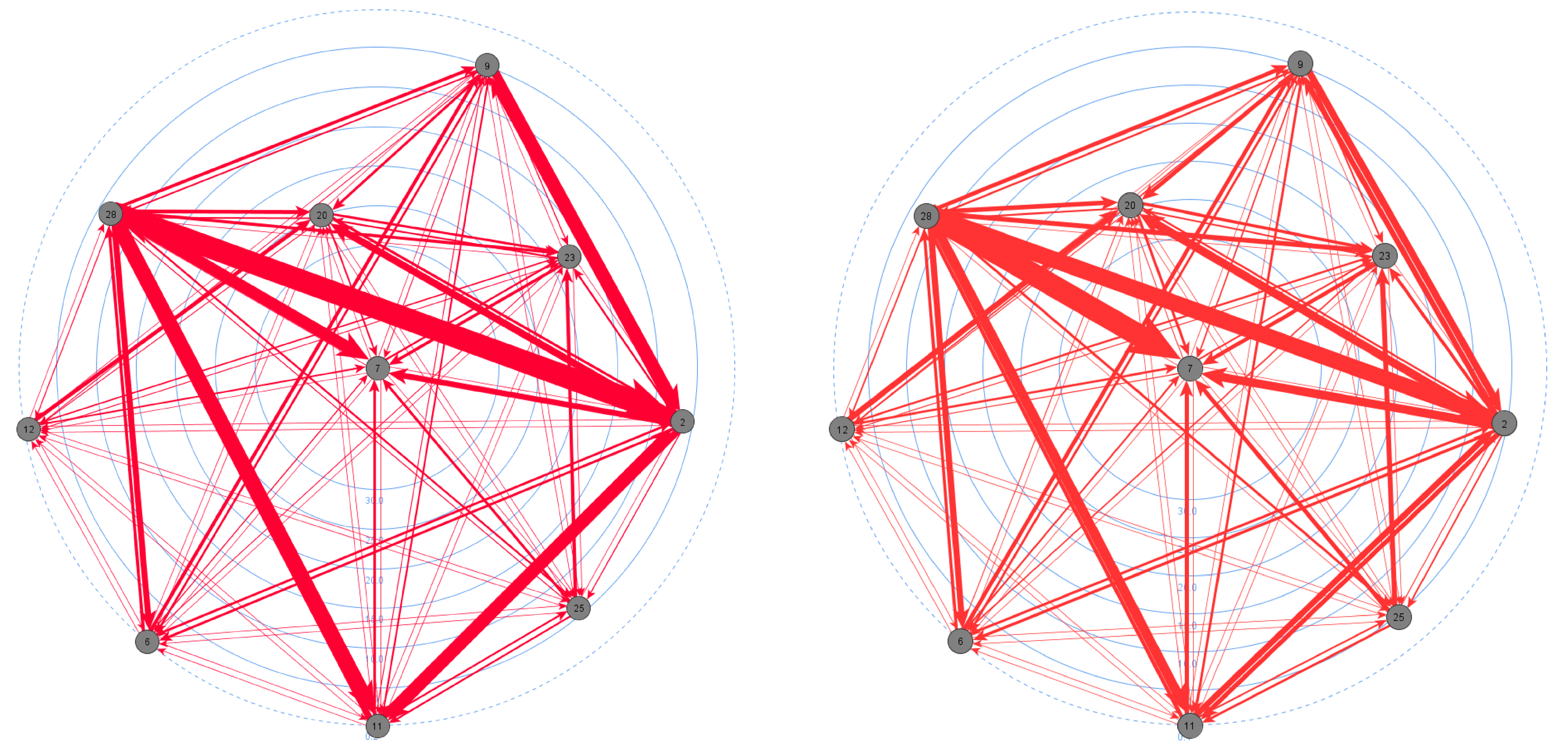
The agreement between the model results and the empirical data suggests that human online conversation behaves as a kind of cascade process, where a short sequence of messages can potentially snowball into a long conversation. If we assume this hypothesis to be true then it may be useful to have a method of "ironing out" these large fluctuations. To achieve this we use the following assumption:

- **The probability that person i sends a message to person j at time t is proportional to the number of messages i has received from j in the previous 8 hours.**

The probability in question is analogous to $f(k_t)$ in our burst model. Now, when a message is sent, we are able to measure how **informative** that message is by the reciprocal of the likelihood of it happening: $1/f(k_t)$. We now define the weight of the edge (i,j) as

$$W_{i,j} = \sum_{m \in S(i,j)} \log\left(\frac{1}{f(k_t(m))}\right) \tag{1}$$

where $S(i,j)$ is the set of all messages from i to j and $k_t(m)$ is the number of messages i has received from j in the 8 hours before $t(m)$, the time that message m was sent.



(a) Weighted by number of messages. **(b)** Weighted with bursts 'ironed out'.

Figure 2: A subnetwork of Twitter users and the edges (mentions) between them. In (2a) the weight of each directed edge is simply the total number of messages sent from one user to another over a one week period (normalised so that the total weight of the network is 1). In (2b) the weight is given by Eq.(1) (then normalised to give a total weight of 1). The weights of many of the edges are similar in both networks; some however, particularly the very heavy edges in (2a), are significantly different owing to the fact that much of that weight comes from bursts.

Ironed networks are steadier over time

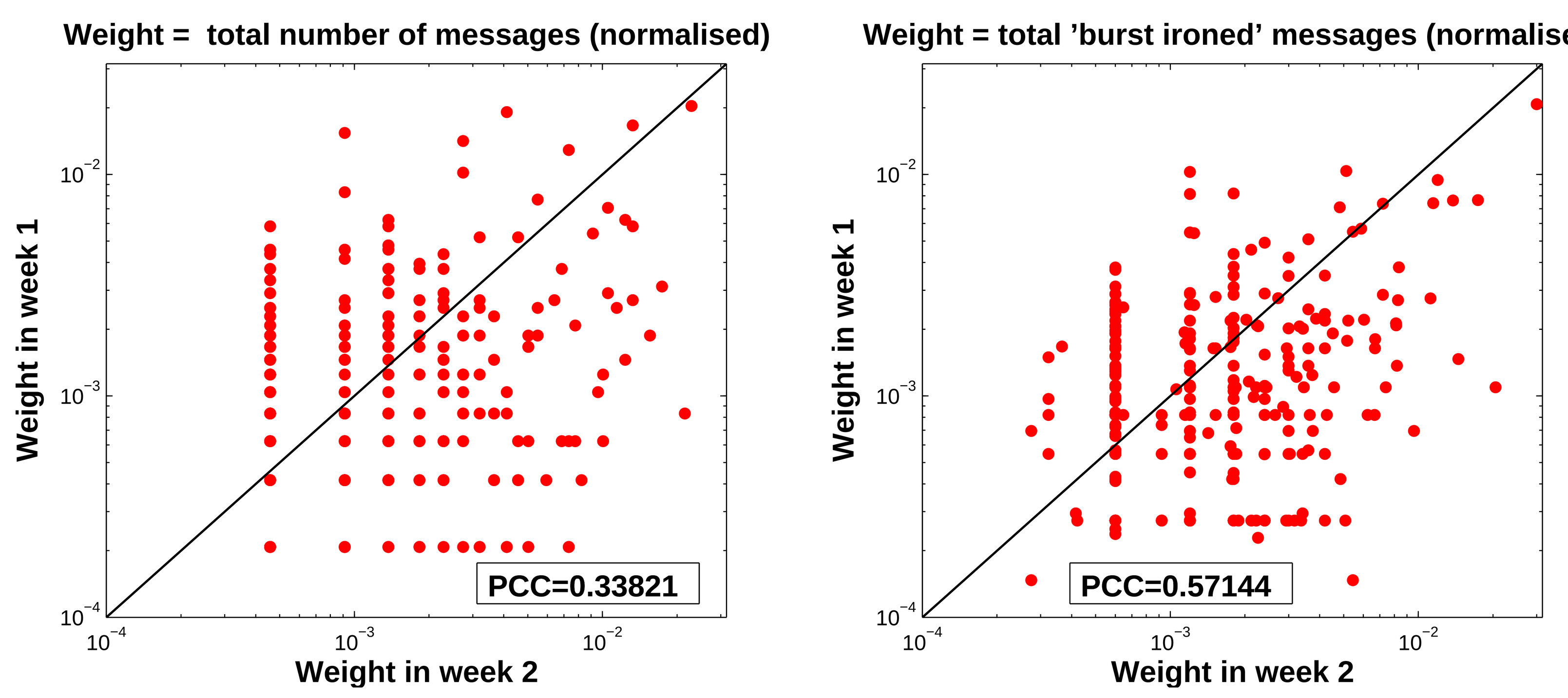


Figure 3: Each red marker represents one edge (i,j) in a 'mentions' network of Twitter users. We look at two consecutive weeks and ask whether the weight of each edge in week 2 is similar to (and therefore can be predicted by) its weight in week 1. Markers which are close to the diagonal line represent edges which did not change much from week 1 to week 2. The two plots show the same empirical network, the only difference being the way the weights of the edges are calculated. We find that burst ironed networks show a higher level of steadiness, quantified by a higher level of correlation between the first week and the second (we have used the Pearson product moment correlation coefficient).

References

[1] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Physical review letters*, 71(24):4083, 1993.

[2] A.-L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[3] E. R. Colman and D. V. Greetham. Memory and burstiness in dynamic networks. *arXiv preprint arXiv:1501.05198*, 2015.

[4] S. Gaito, M. Zignani, G. P. Rossi, A. Sala, X. Zhao, H. Zheng, and B. Y. Zhao. On the bursty evolution of online social networks. In *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, pages 1–8. ACM, 2012.

[5] K.-I. Goh and A.-L. Barabási. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4):48002, 2008.

[6] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész. Universal features of correlated bursty behaviour. *Scientific reports*, 2, 2012.

[7] M. Karsai, N. Perra, and A. Vespignani. Time varying networks and the weakness of strong ties. *Scientific reports*, 4, 2014.

[8] T. Maillart, D. Sornette, S. Frei, T. Duebendorfer, and A. Saichev. Quantification of deviations from rationality with heavy tails in human dynamics. *Phys. Rev. E*, 83:056101, May 2011.

[9] A. Moinet, M. Starnini, and R. Pastor-Satorras. Burstiness and aging in social temporal networks. *arXiv preprint arXiv:1412.0587*, 2014.

[10] D. Peng, X.-P. Han, Z.-W. Wei, and B.-H. Wang. Punctuated equilibrium dynamics in human communications. *arXiv preprint arXiv:1410.5158*, 2014.

[11] S. Vajna, B. Tóth, and J. Kertész. Modelling bursty time series. *New Journal of Physics*, 15(10):103023, 2013.

[12] A. Vazquez. Impact of memory on human dynamics. *Physica A: Statistical Mechanics and its Applications*, 373(0):747 – 752, 2007.

[13] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(3):036127, 2006.

[14] C. L. Vestergaard, M. Génois, and A. Barrat. How memory generates heterogeneous dynamics in temporal networks. *Physical Review E*, 90(4):042805, 2014.

Acknowledgments

This work is funded by the RCUK Digital Economy programme via EPSRC Grant EP/G065802/1 'The Horizon Hub'.