Statistical characterisers of transport on networks

Neelima Gupte
Dept. of Physics
IIT Madras,
Chennai, 600036.

Collaborators: Satyam Mukherjee, Anirudh Vij

Networks

The study of networks has been a topic of vigorous recent interest.

- A network consists of assemblies of elements, and can be represented by nodes plus links between nodes.
- Each node may be capable of some function and may have some capacity.
- Thus the network is capable of carrying out some task, or of supporting some dynamical processes.
- Networks are ubiquitous in the real world in both natural and engineered contexts.

Networks: Examples

Both natural and engineered networks are seen.

- Power grids, Internet, Traffic networks, Telephone networks
- Metabolic networks, neural networks, ecological networks, food-webs.
- Collaborative networks, friendship networks, co-worker networks.

Transport on networks

- Transport processes on metabolic and other biological networks
- Traffic of information packets on computer or communication networks
- Road traffic in transportation networks, air traffic in airport networks

The efficiency and optimisation of transport on the networks is controlled by

- the structure and topology of the network
- the mechanism of transport

Communication network Models

- Scale-free networks.
- Can be based on regular $2-d$ geometries and incorporate geographic separations.
- Networks of hosts and routers (Sole and Valverde)
- Incorporate clustering and hubs (Rosenfeld).

Connections and routing

- Gradient flows on scale-free geometries (Bassler and Toroczkai)
- Distance based connections (Waxman)
- Random assortative connections (Singh and Gupte)
- A variety of routing algorithms are possible.
- Important to note that these networks can reproduce some of the characterisitics of realistic internet traffic.

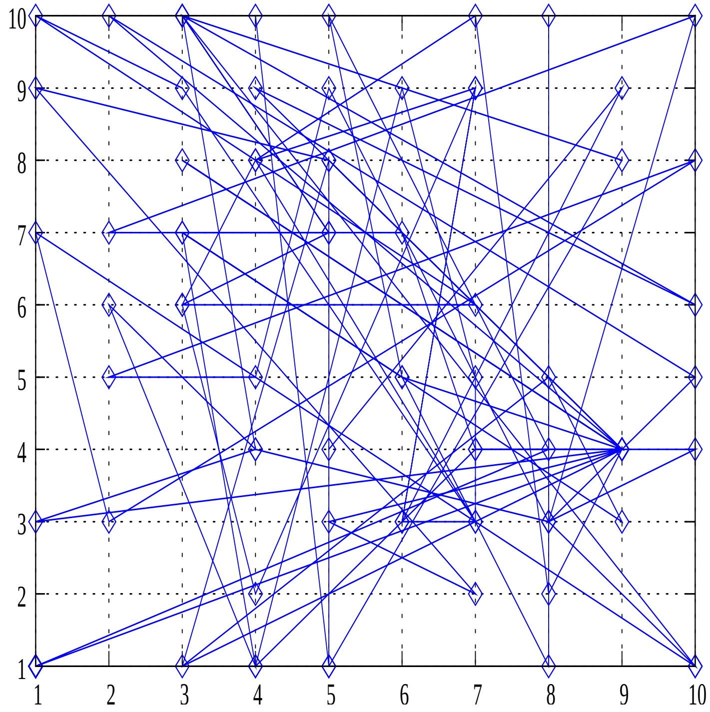
Models under discussion

- Communication networks based on two-d and 1-d lattices. The models incorporate local clustering and geographic separations.
- Despite the regular geometry, traffic on 2-d networks reproduces the characteristics of realistic internet traffic (Sawada and Ohira, Lawniczek).
- Single message transfer and multiple message transfer are studied.
- Under multiple message traffic, a transition from a decongested phase where traffic flows freely, to a congested phase where traffic jams is seen.

Models under discussion

- Statistical characterisers: Average travel times, travel time distributions, waiting time distributions.
- These statistical quantities show characteristic signatures of congestion or decongestion, and the network topology.
- Synchronisation (both complete synchronisation, and phase synchronisation) is seen between the queues at the most frequented hubs in the congested phase. Synchronisation is lost as the queues clear.
- A synchronisation to desynchronisation transition in seen in the queue lengths at the most frequently visited hubs for the models.
- Real-life networks show these effects. We demonstrate these for the air-port network of the U.S. and the IITM campus network.

The Waxman network



The Waxman network.

The Waxman network

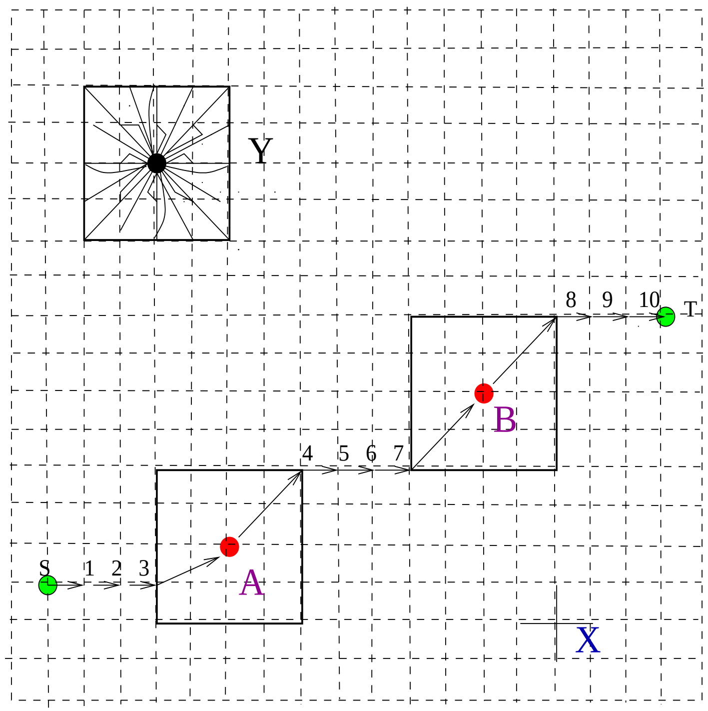
- The Waxman graphs are generated on a co-ordinate grid with the probability of a link betweennode a and node b being given by$d$ being given by

$$P(a,b) = \beta \exp(-\frac{d}{\alpha M}) \quad (0)$$

  where $0 < \alpha, \beta < 1$, d is the Euclidean distance between a and b, and $M = \sqrt{2}L$.

- Larger values of β correspond to larger link densities, and smaller values of α increase the density of shorter links as compared to longer ones.
- The Waxman networks are popular models of intranet topologies.

The clustered network

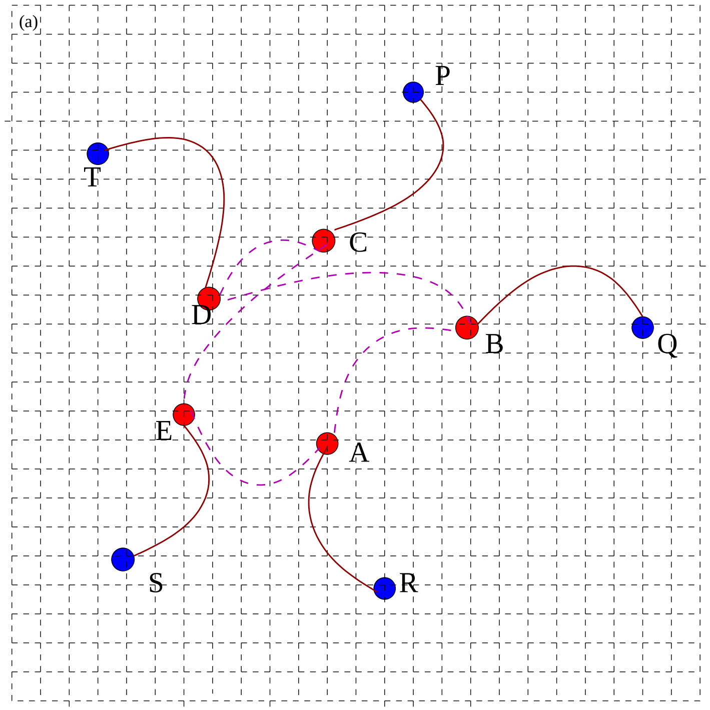


A regular $2-d$ lattice. Here, X is an ordinary node, Y is a hub and the dotted square shows the area of influence. A typical path is shown.

Single Message Transfer

- Any node can function as a source or target node for a message and can also be a temporary message holder or router.
- The metric distance between any pair of source (is, js) and target (it, jt) nodes on the network is defined to be the Manhattan distance $D_{st} = |is - it| + |js - jt|$.
- The message transfer between source and target takes place from node to node via the shortest path utilising the hubs.
- The constituent nodes of the hub transfer the message directly to the hub.
- The hub transfers messages to the peripheral node nearest the target.

Connecting the hubs: Random assortative connections



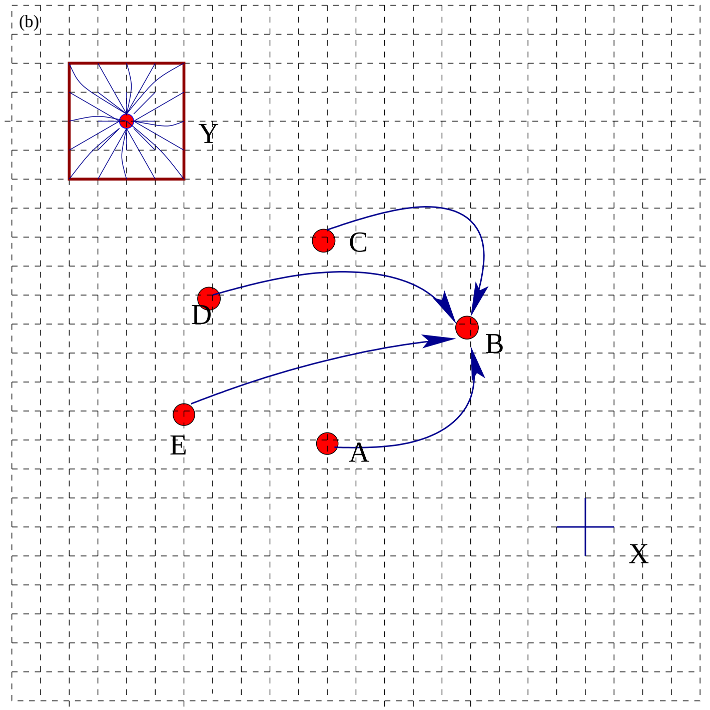
Random assortative connections between hubs. Speed up message transfer and achieve decongestion.

Connecting the hubs

Message transfer can be speeded up by setting up hub to hub connections.

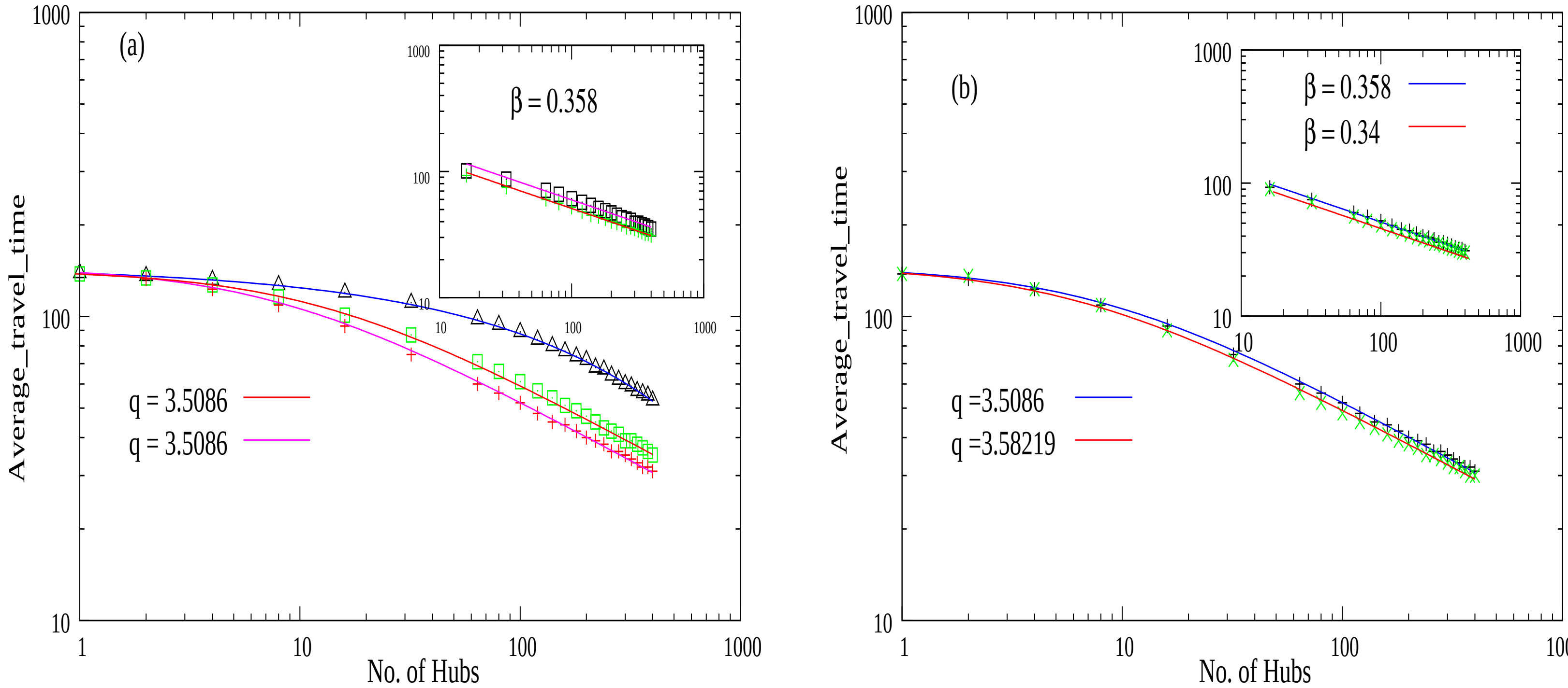
- Hub Capacity:
  This is defined to be the number of messages the hub can process simultaneously.
- Gradient connections:
  Each hub is randomly assigned some message capacity between one and C_{max}. A gradient connection is assigned from each hub of capacity less than C_{max} to all the hubs with the maximum capacity (C_{max}).
- Random Assortative connections:
  Assortative connections one way, or two way, are made from each hub to two randomly chosen other hubs. Here, the hub capacities are all unit.

Connecting the hubs: Gradient connections



Gradient connections between hubs. These can shortcut the message transfer.

Average travel times

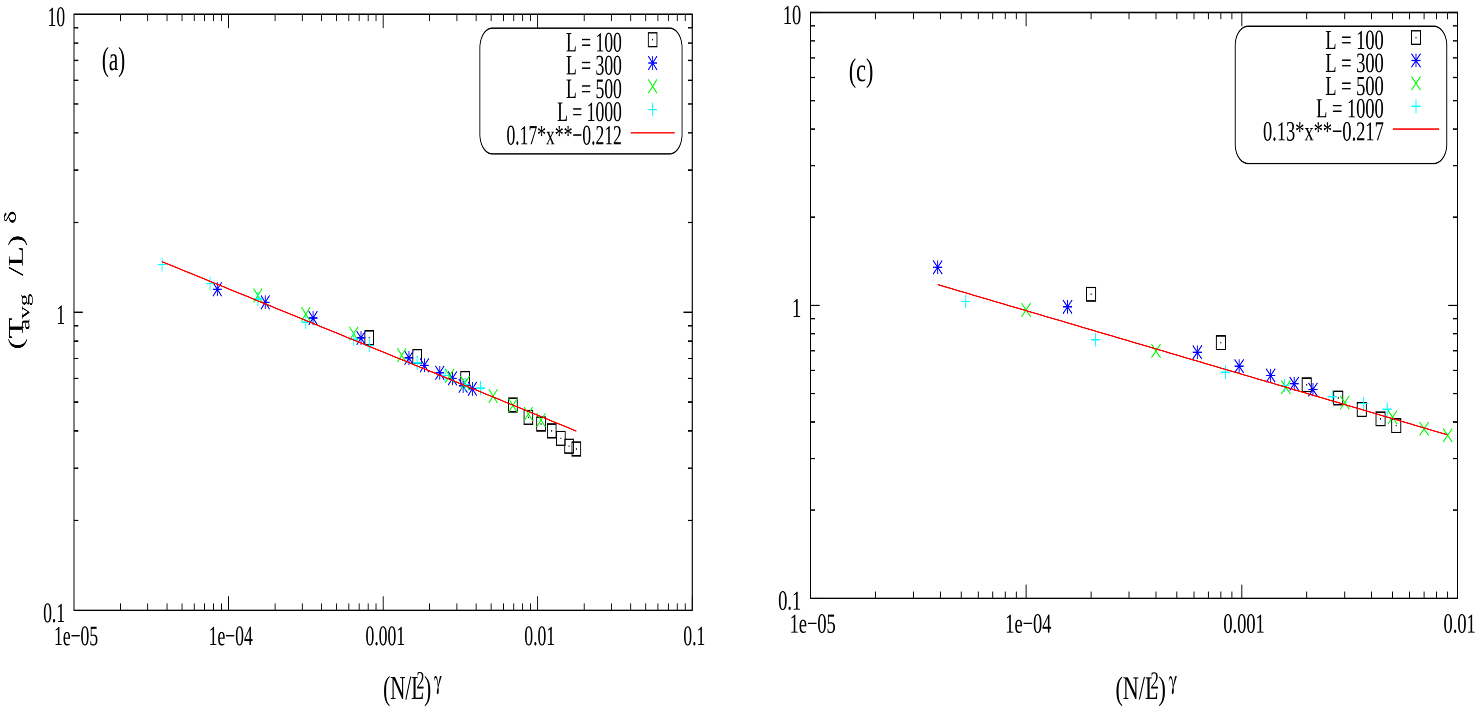


Baseline data (Δ) $f(x) = Qexp[-Ax^{\alpha}]$; gradient data (□), one-way assortative (+), two-way assortative ($times$) $f(x) = A(1-(1-q)x/x_0)^{(1/(1-q)}$).

Average travel times

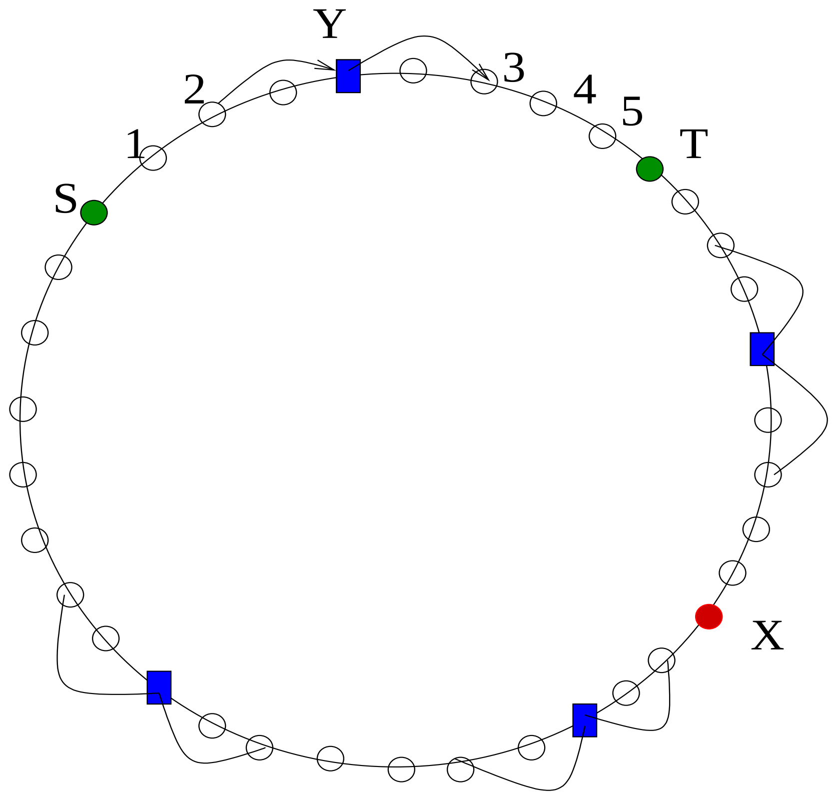
- The average travel time t_{avg} for messages shows stretched exponential behaviour as a function of hub-density on the baseline. Here, $f(x) = Qexp[-Ax^{\alpha}]$, where α = 0.50 $\pm$ 0.011, $A = 0.051$ and $Q = 146$.
- However, the gradient data fits a q-exponential $f(x) = A(1 - (1 - q)x/x_0)^{(1/(1-q)}$with $q = 3.51$, $A = 142$ and $x_0 = 0.03$.
- The one-way assortative connections and two way assortative connections are also $q-$exponential functions.
- The tails of the $q-$ exponentials are power-laws. Thus average travel time falls rapidly at high hub density.

Finite Size Scaling

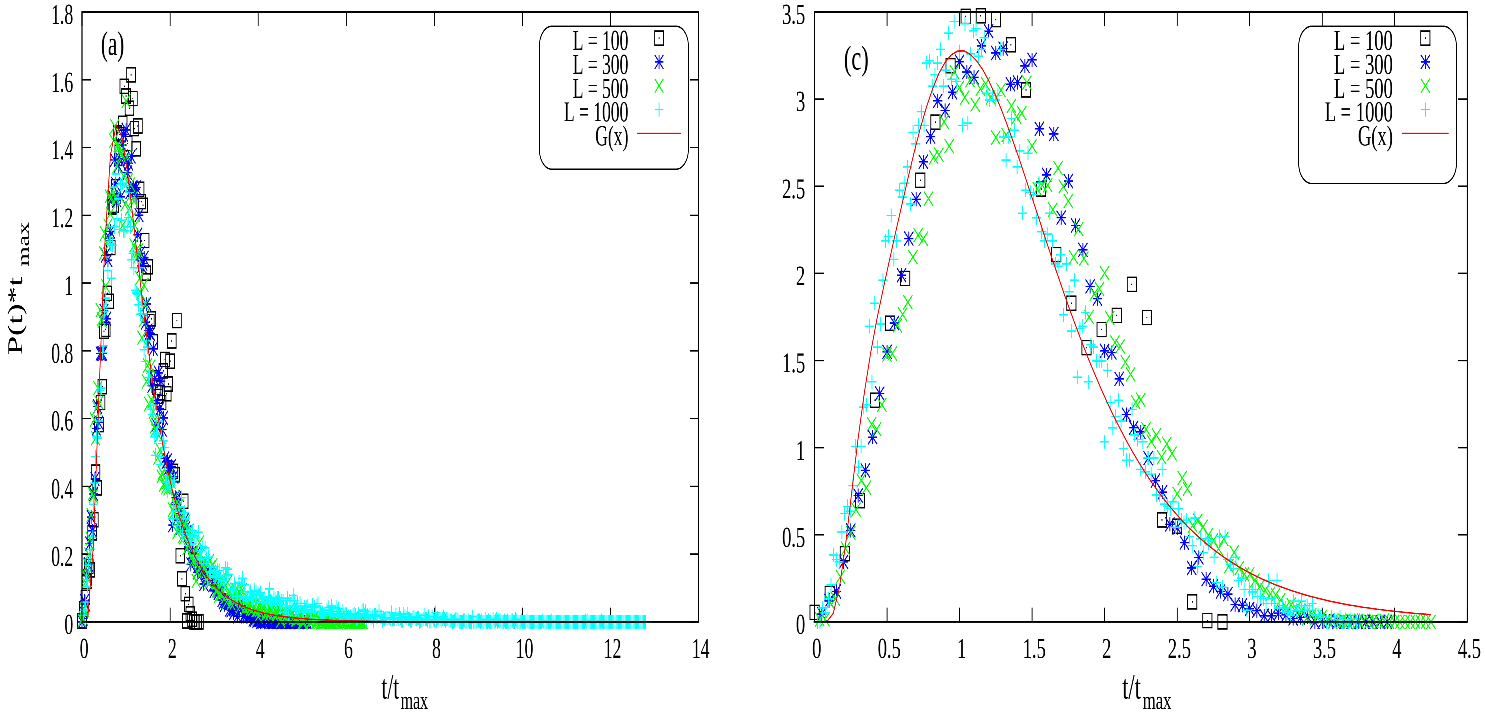


Data collapse for different $L-$s. The final curve fits a power law of with exponent β = 0.212 (gradient) and β = 0.217 (one way assort).

A $1-d$ *network*



Travel time distribution for single messages

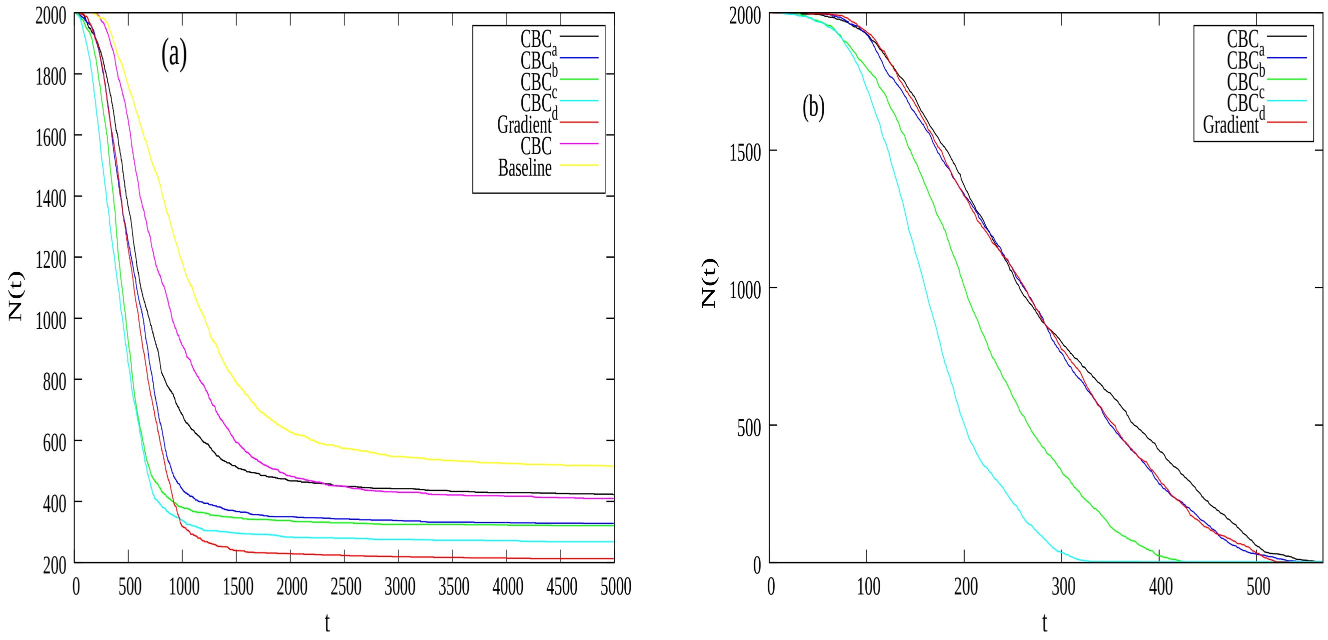


The scaled travel time distribution for (a) the Gradient mechanism,(b) the one way assortative mechanism.

Congestion and decongestion of traffic

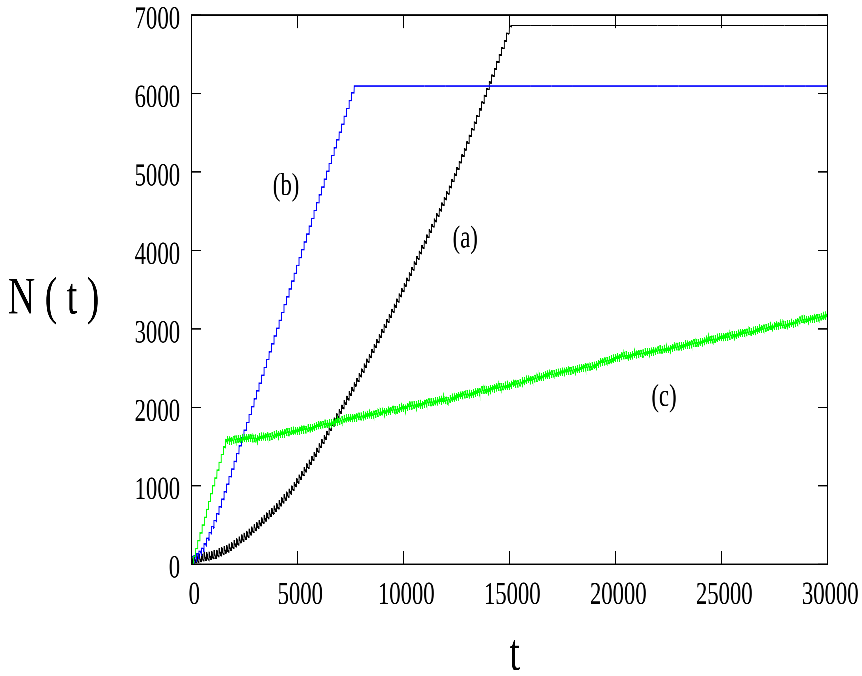
- Realistic networks experience congestion problems under multiple message transfer due to capacity limitations.
- Hubs which see heavy traffic are prone to trap messages.
- Such hubs identified by defining a co-efficient of betweenness centrality $CBC = \frac{N_k}{N}$ where N_k is the number of of hubs through a given hub k and N is the total number of messages running through the lattice.
- Signatures of congestion can be seen in statistical characterisers.

Congestion and trap formation: One time deposition



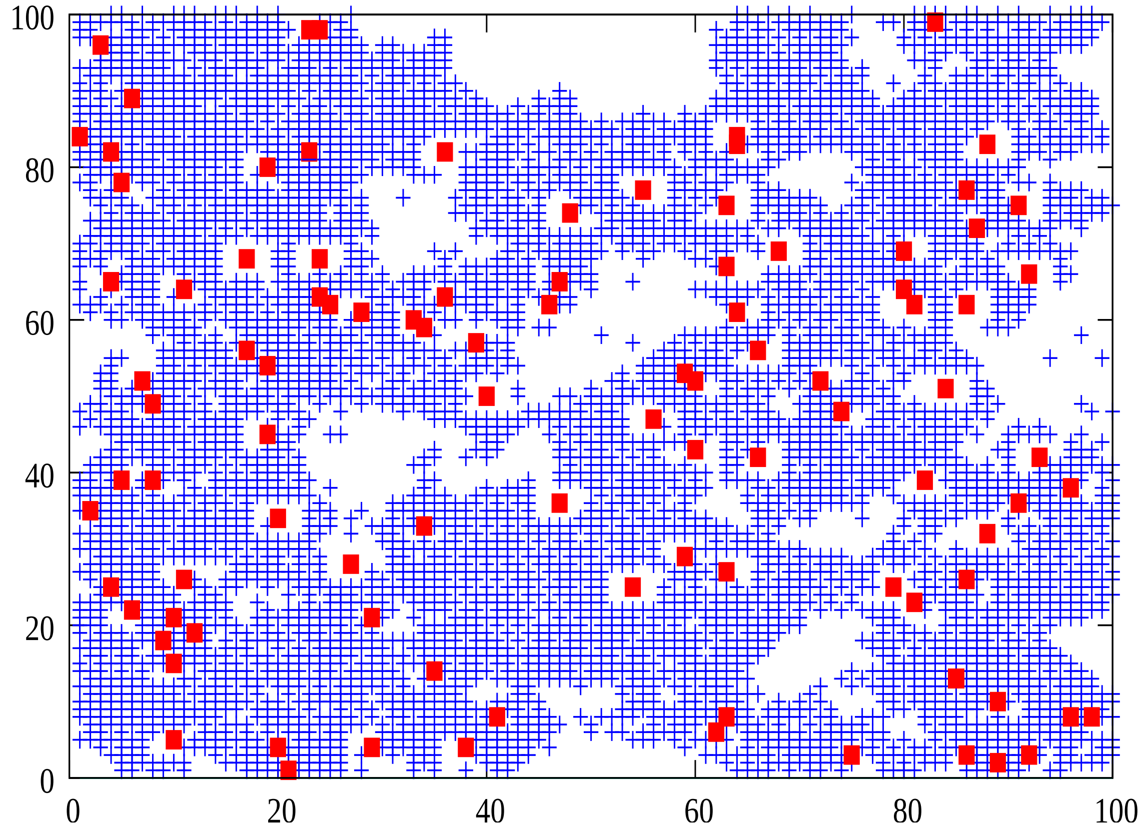
(a) 50 hubs and (b) 400 hubs in a 100×100 latticewith D_{st} = 142. (b) assortative and gradient scheme. Initially $N = 2000$ messages are released.

Congestion and trap formation: Constant density traffic



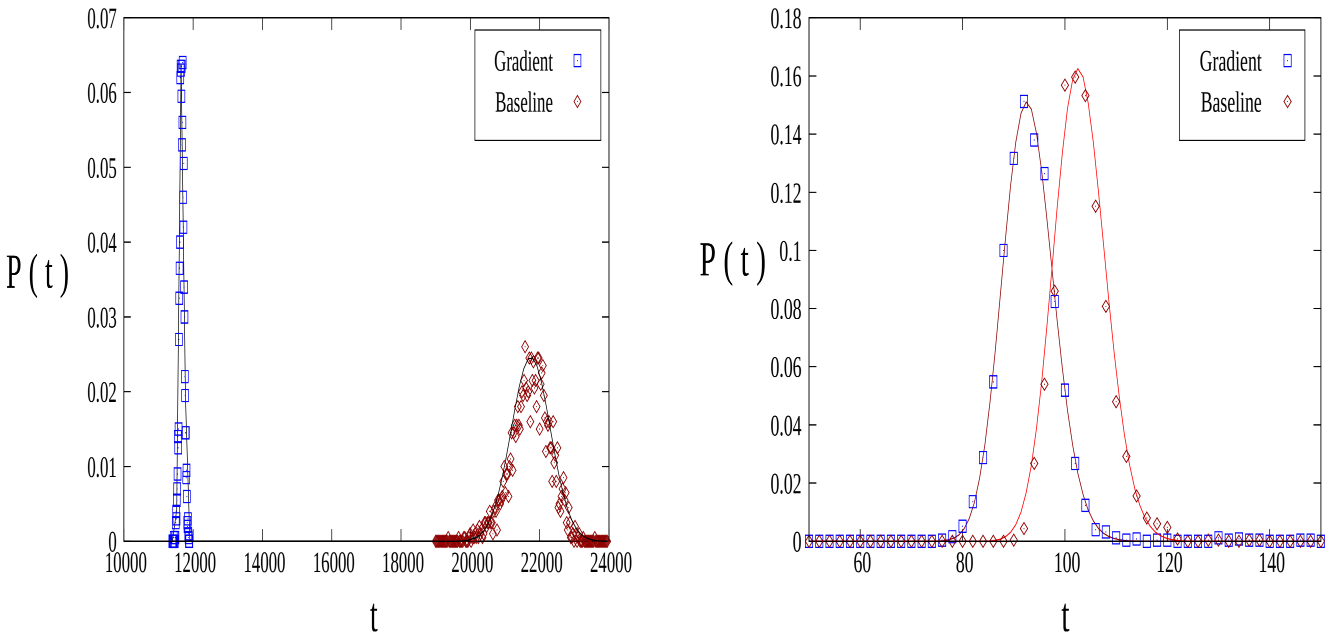
The plot of number of messages $N(t)$ flowing on the lattice as a function of time t for (a) the Waxman topology network and the baseline mechanism for (b) the locally clustered $2-d$ lattice and (c) the $1-d$ ring network.

Trap formation: Constant density traffic



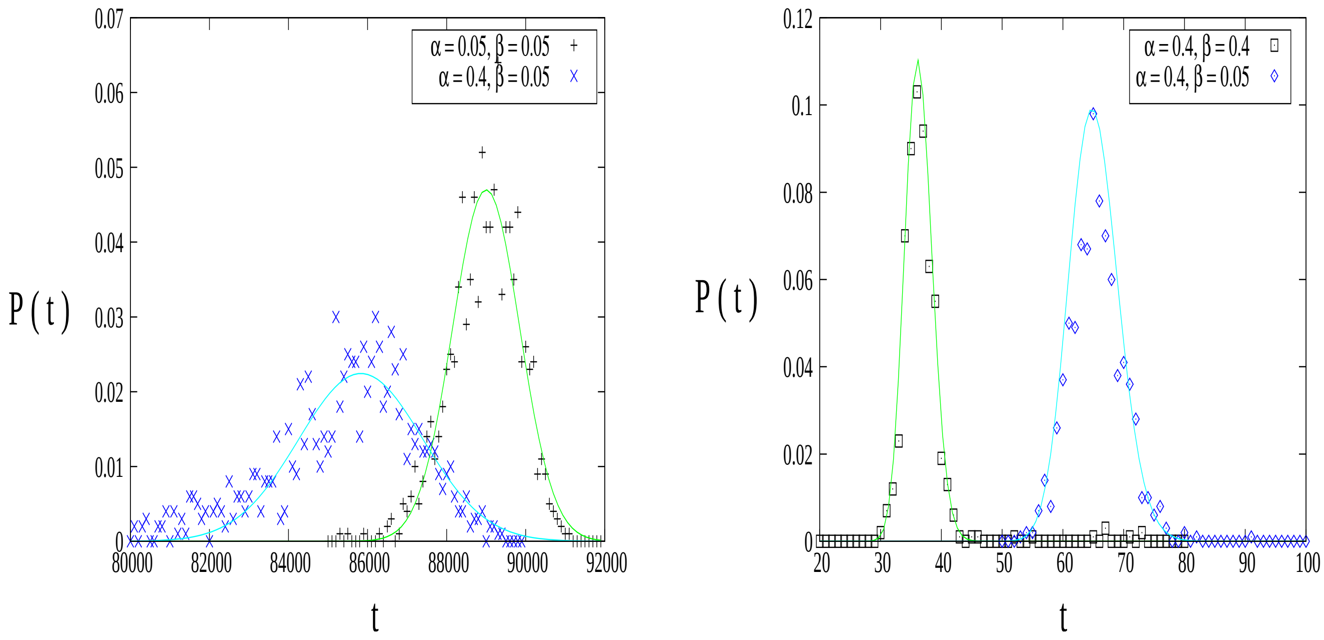
Trap formation for the baseline network

Travel time distributions: gradient and baseline



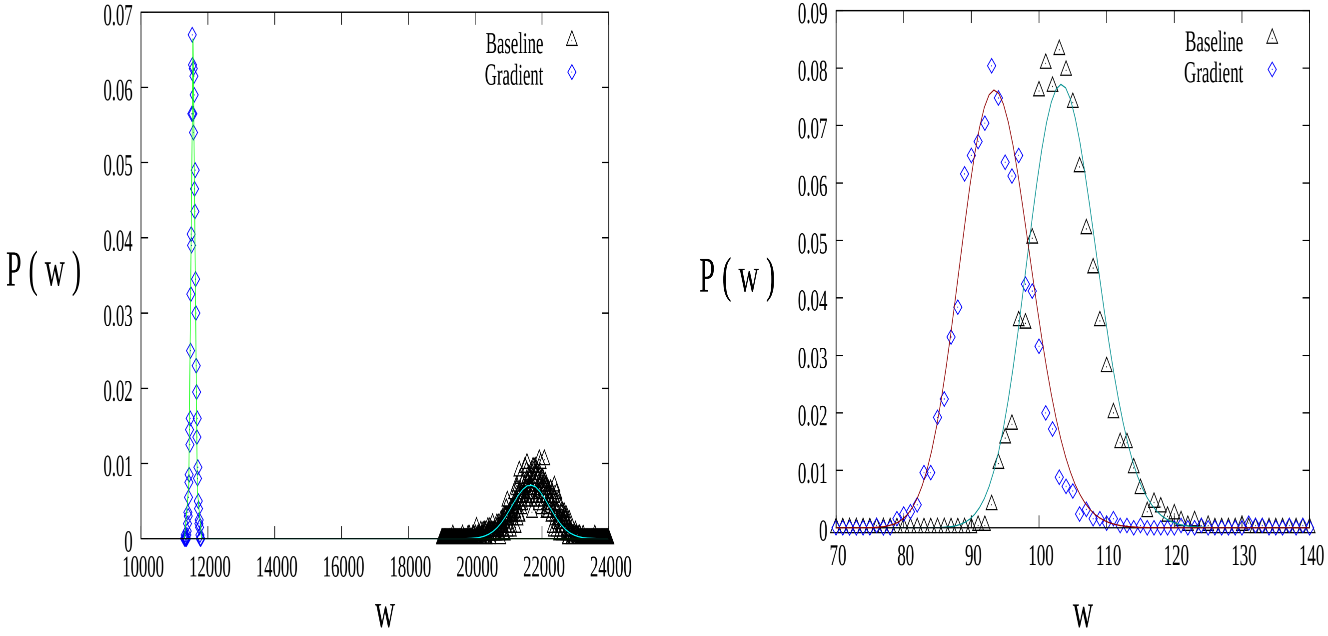
(a) Gaussian σ (gradient) = 64.07 ($\chi^2 = 0.026$) and σ (baseline) = 567.31 ($\chi^2 = 0.195$) σ (gradient) = 0.05 ($\chi^2 = 0.04$) and (b) log normal σ (baseline) = 0.056 ($\chi^2 = 0.04$).

Travel time distributions: Waxman networks



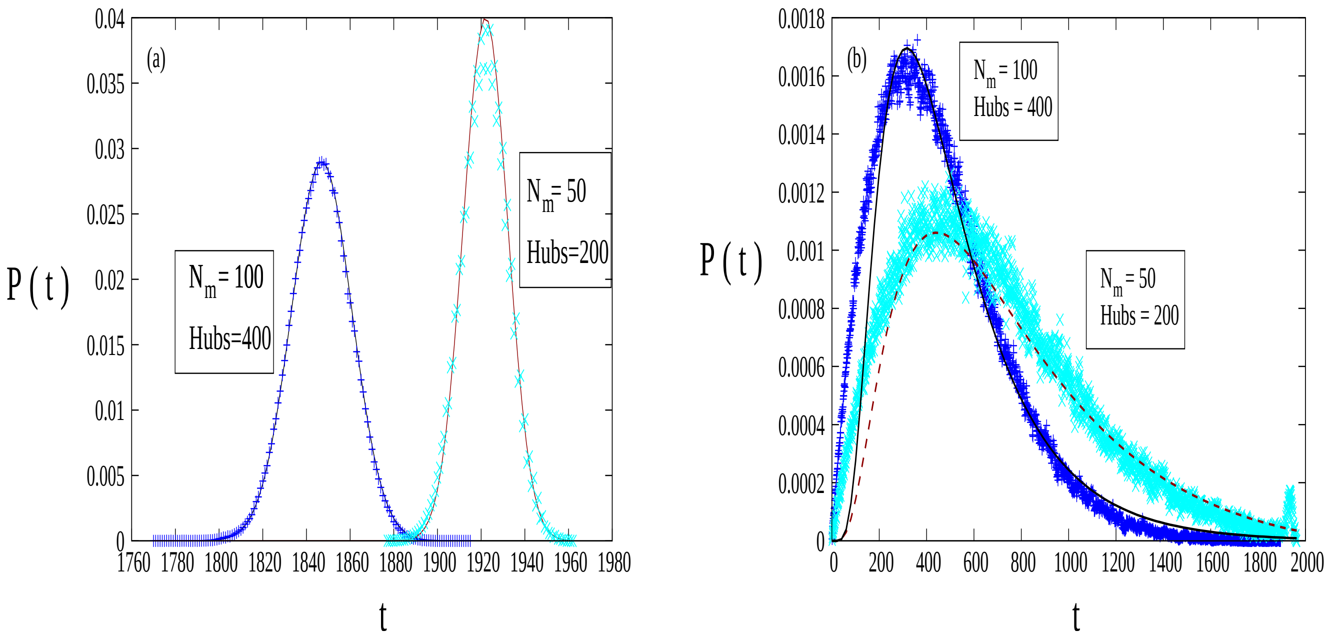
(a) Gaussian σ ($\alpha = 0.05, \beta = 0.05$) = 850.37 ($\chi^2 = 0.463$) and σ ($\alpha = 0.4, \beta = 0.05$) = 1576 ($\chi^2 = 0.852$, (b) log-normal σ ($\alpha = 0.4, \beta = 0.4$) = and 0.062 ($\chi^2 = 0.06$) and σ ($\alpha = 0.4, \beta = 0.05$) = 0.06

Waiting time distributions: gradient and baseline



(a) Gaussian σ (gradient) = 60 ($\chi^2 = 0.02$) and σ (baseline) = 560 ($\chi^2 = 0.2$); (b) log-normal σ (gradient) = 60 ($\chi^2 = 0.02$) and σ (baseline) = 560 ($\chi^2 = 0.2$).

Travel time distributions: $1-d$ ***ring***

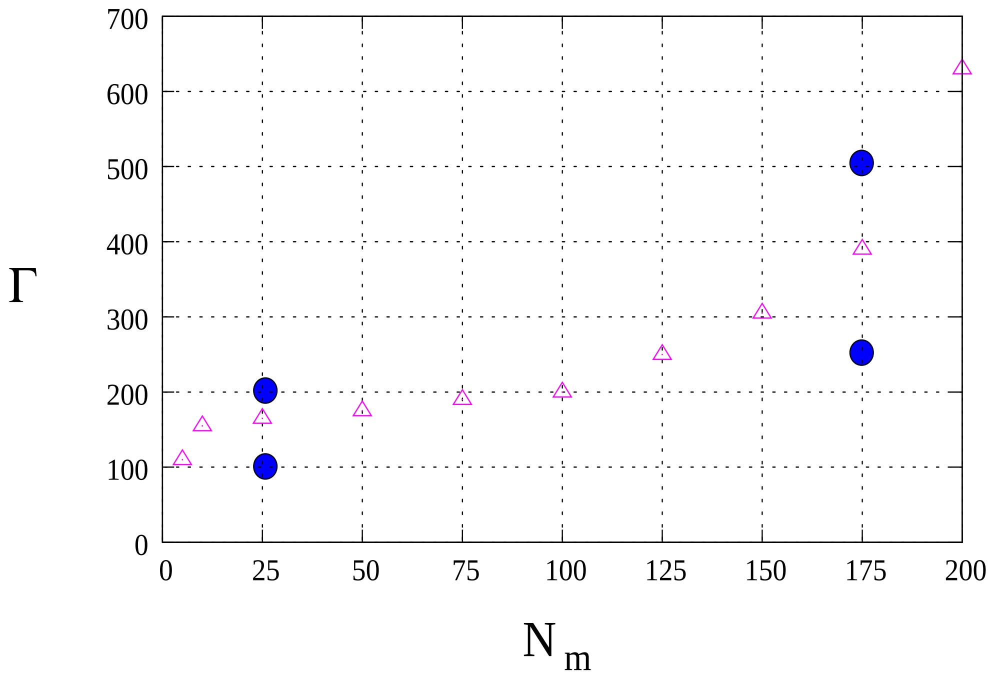


$1-d$ ring. (a) Gaussian, σ is (i) 13.86 ($\chi^2 = 0.0086$) for $N_m = 100$ and 400 hubs (ii) 10.21 ($\chi^2 = 0.0095$) for $N_m = 50$ and 200 hubs. (b) Log-normal with a power law correction (i) $N_m = 100$ and 400 hubs, $\sigma = 1.42$ ($\chi^2 = 0.14$), $\beta = 0.88$, $B = -0.0009$ and (i) $N_m = 50$ and 200 hubs, $\sigma = 1.79$ ($\chi^2 = 1.75$),

Waiting time distributions

- The distribution of waiting times is normal in the congested phase
$$\frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(w-a)^2}{2\sigma^2})$$
and log-normal in the decongested phase
$$\frac{1}{w\sigma\sqrt{2\pi}} exp(-\frac{(\ln w-\mu)^2}{2\sigma^2}).$$
- For the $1-d$ ring we have the distribution
$$\frac{1}{t\sigma\sqrt{2\pi}} \exp(-\frac{(\ln t-\mu)^2}{2\sigma^2})(1+Bt^{-\delta})$$

Phase Diagram



Here N_m is the number of messages and Γ is the posting rate.

Phase Diagram

Values of Γ_c for different values of N_m (baseline)

N_m	Γ_c
5	110
10	155
25	165
50	175
75	190
100	200
125	250
150	305
175	390
200	630

The stochastic process for transport

- The displacement of a message on the network depends on factors that are partly systematic, and partly random, with the random element arising due to interference from other messages and limitations of hub capacity.
- Hence, message transfer in both the congested and decongested phases can be modelled by stochastic differential equations.

The stochastic process for transport

In the congested phase, the displacement X_t of the message at a time$t$ can be modelled by the equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

here, $\mu(X_t)$ represents the drift co-efficient, $\sigma(X_t)$ is the diffusion co-efficient and W_t is a Wiener process.
The probability distribution $f(X,t)$ satisfies the forward Kolmogorov equation

$$\frac{\partial f}{\partial t} = \frac{(\partial \mu(X) f(X,t))}{\partial X} + \frac{\partial^2}{\partial X^2}(\sigma^2(X) f(X,t))$$

The stochastic process for transport

The stationary solution, i.e. $\frac{\partial f}{\partial t} = 0$ can be found using Wright's formula

$$f(x) = \frac{N}{\sigma^2} \int_{-\infty}^{x} \frac{\mu(s)}{\sigma^2(s)} ds$$

where N is a normalisation constant.
If the drift co-efficient is of the form $\mu(X_t) = (\bar{\mu} - X_t)$ and the diffusion co-efficient $\sigma(X_t)$ is a constant,then, the stationary probability distribution $f(X)$ turns out be of the normal form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp - \frac{(X - \bar{\mu})^2}{2\sigma^2}$$

The stochastic process for transport

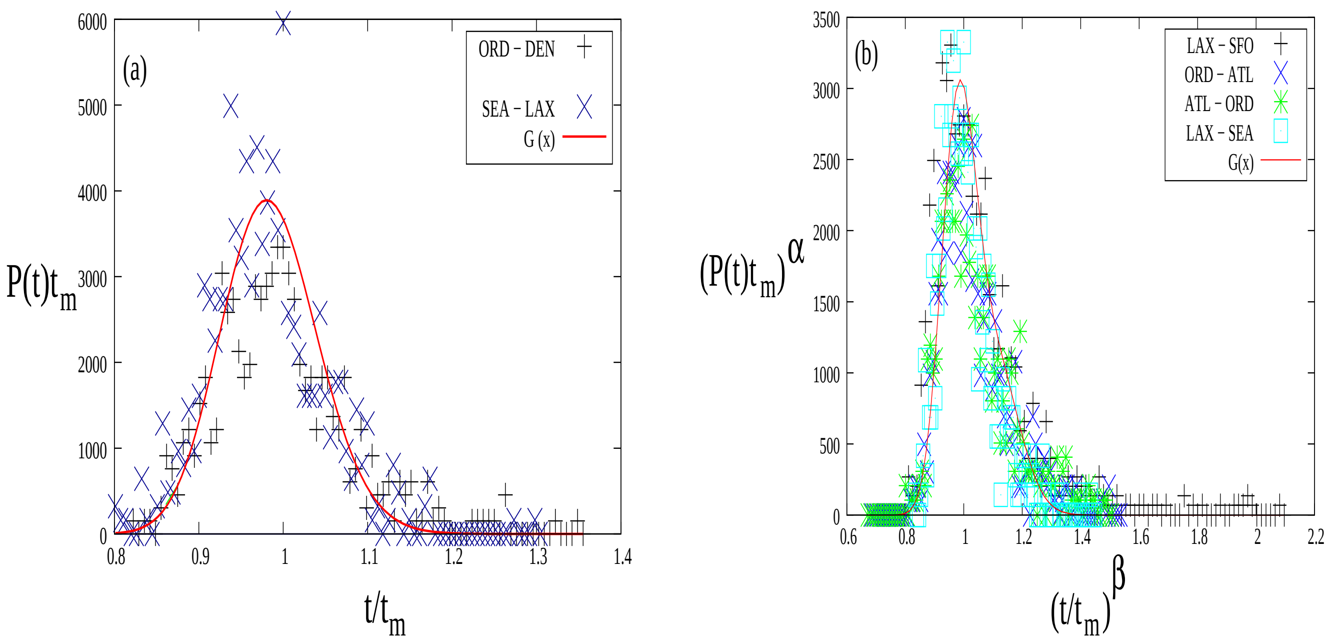
On the other hand, in the decongested phase, the process can be modelled by the equation

$$dX_t = \mu(X_t)X_t dt + \sigma X_t dW_t$$

Let $S(X)$ = $\log(X)$. Then the form of the stichastic equation in the decongested phase can be reduced to the form in the congested phase. Then, using $f(X) = f(S)\frac{dS}{dX}$, the form of $f(X)$ in the decongested phase can be found to be

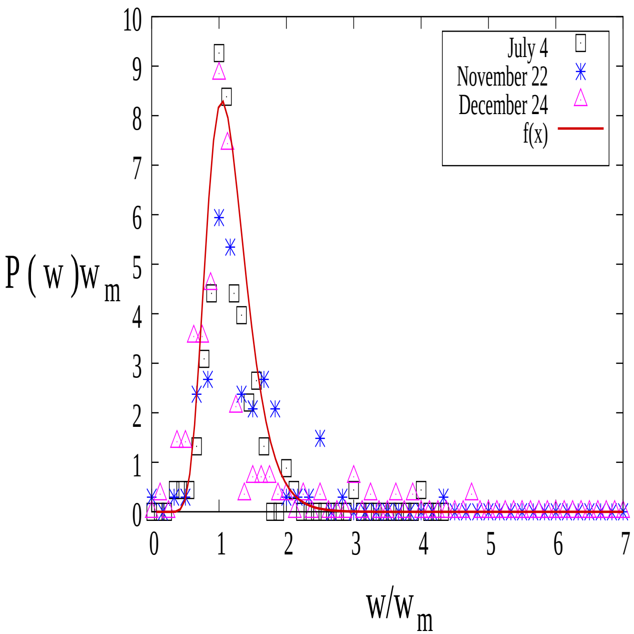
$$f(X) = \frac{1}{\sigma X\sqrt{2\pi}} exp - \frac{(\ln X - \bar{\mu})^2}{2\sigma^2}$$

Airport traffic

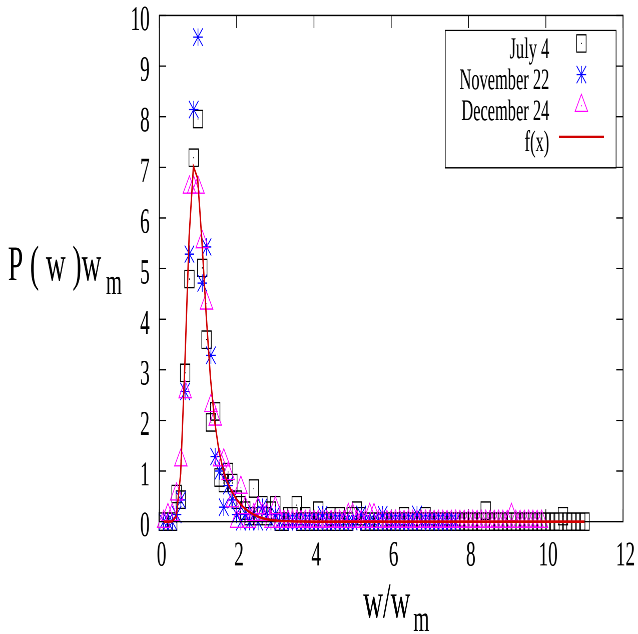


Airport traffic data for the airports Chicago (ORD), Atlanta (ATL), Seattle (SEA), Los Angeles (LAX), Denver (DEN), $\alpha = 0.95$, $\beta = 1.01$

Airport traffic: Delay times



(a)



(b)

(a) The scaled delay time distribution of flights of UA arriving at the airport SFO show log-normal behavior. (b) For the flights of AA arriving at DFW, the distribution fits into a log-normal with a power law correction. The data is plotted for July 4 2007, November 22 2007 and December 24 2007 .

Airport traffic: Non-flight parameters

Name	Pass. (mil)	Aircr. (ths)	Cargo (KTons)	RW	HP	TP.	TC	Area (K.ac
ORD	77.0	958.6	2003	7	1	4	15	7.6
DEN	47.3	598.4	645.4	6	0	5	4	33.4
DFW	60.2	699.7	797.3	7	1	5	16	18.1
ATL	84.8	976.4	746.5	5	1	2	2	4.7
LAX	61.0	656.8	1907.5	4	1	9	4	3.5
MIA	32.5	384.5	1830.6	4	0	9	24	3.3
SFO	34.9	379.5	752.1	4	0	3	11	5.2
SEA	29.4	317.9	641.7	3	0	16	3	2.5
JFK	43.7	378.4	1636.4	4	4	10	35	5.2

Self organised map for clustering

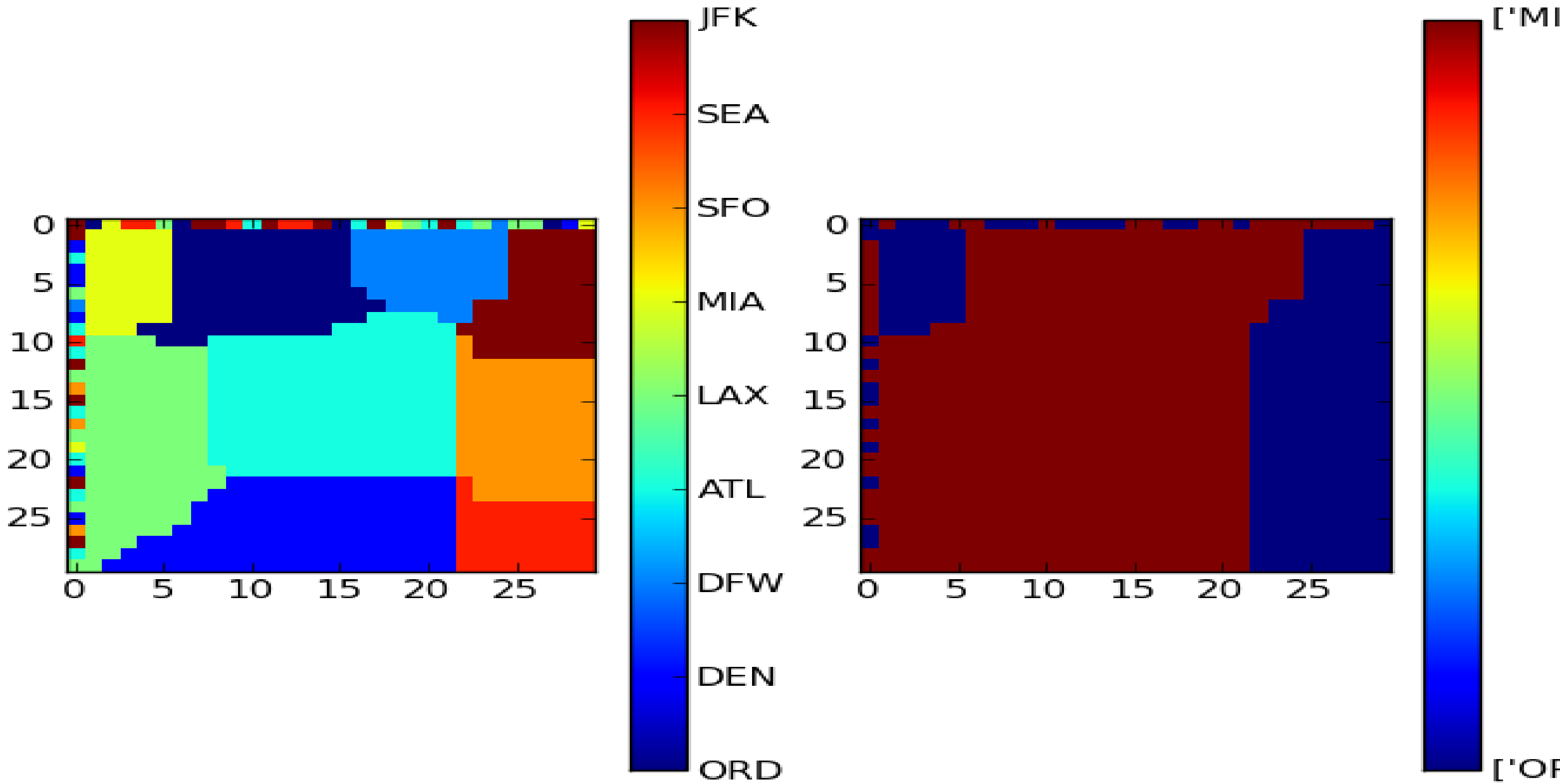
An SOM was used to cluster the air-ports using non-flight data.

- An SOM based on a $2-d$ grid was used to cluster the air-ports using non-flight data.
- Initial vectors were W_i were random
- The neighbourhood function $\theta(t)$ was a truncated Gaussian.
- Input vectors $D(j)$ were the non-flight parameters of the air-ports.

Self organised map for clustering

- Update $W_i(t+1) = W_i + \alpha\theta(t)(D(j) - W_i)$, for the best matching unit.
- After iterations, output vectors assigned based on distance between W_i and input vectors.
- Finally only two parameters turned to be crucial, aircraft movements and runways.

Result of SOM



A run of the SOM for 2 parameters. Raw result on left classification into Class I and II

Classification

1. Class I
   (a) Atlanta (ATL)
   (b) Chicago (ORD)
   (c) Dallas (DFW)
   (d) Denver (DEN)
2. Class II
   (a) Los Angeles (LAX)
   (b) Miami (MIA)
   (c) New York (JFK)
   (d) San Francisco (SFO)
   (e) Seattle (SEA)

Synchronisation

- In the congested phase, the queue lengths for some pairs from the hubs of show phase synchronization and complete synchronization as a function of time.
- A cascading master-slave relation is seen between the hubs, with the hubs of high CBC driving the lower ones.
- The queue lengths are seen to synchronize during the congested phase. In the decongested phase the queues desynchronise.

Queue lengths and synchronisation

- The queue at a given hub is defined to be the number of messages which have the hub as a temporary target.
- Two queue lengths $q_i(t)$ and $q_j(t)$ are said to be completely synchronized if

$$q_i(t) = q_j(t)$$

  where $q_i(t)$ is the queue at the i^{th} hub. Complete synchronisation is seen for certain pairs of hubs with random assortative connections.

Phase Synchronisation

- The phase at a given hub is defined as

$$\Phi_i(t) = tan^{-1} \frac{q_i(t)}{< q_i(t) >}$$

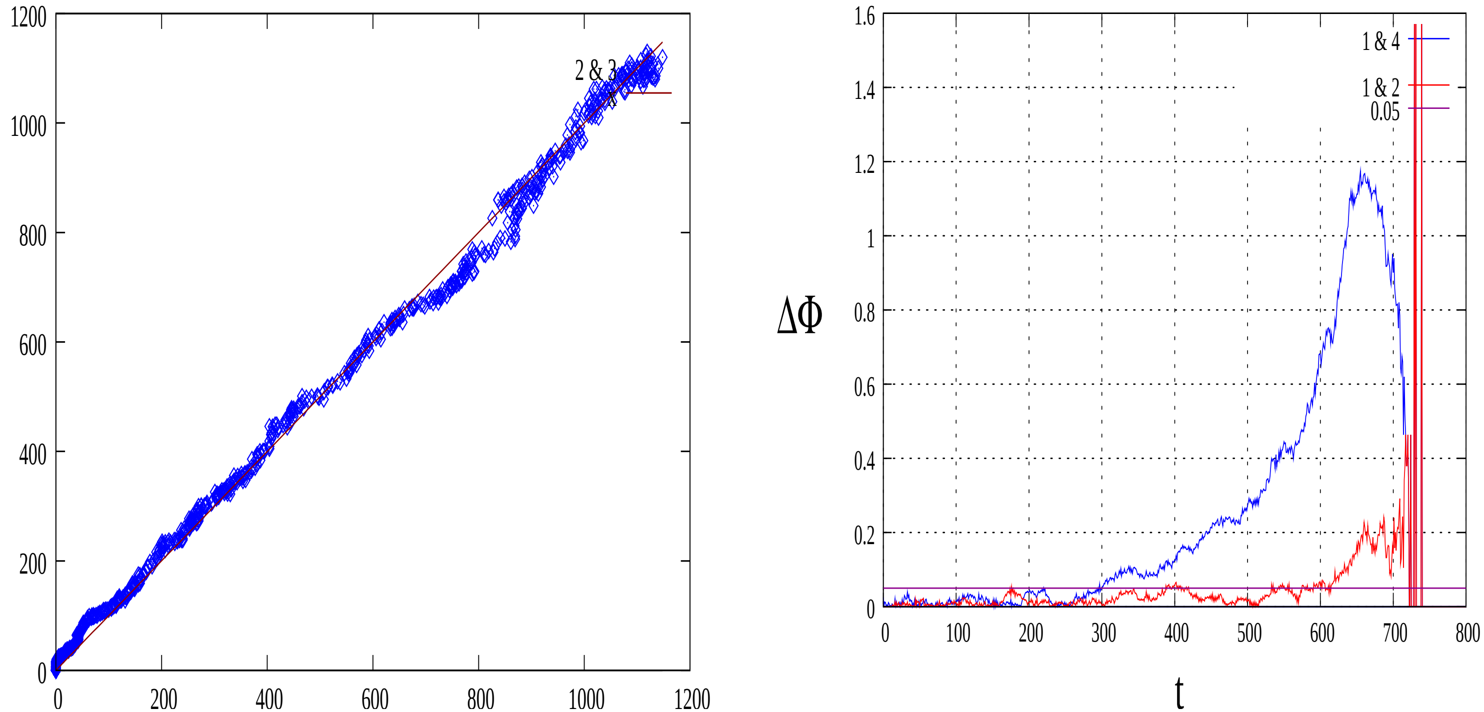
- The queue lengths are phase synchronized if

$$|\Phi_i(t) - \Phi_j(t)| < Const$$

  where Φ_i(t) and Φ_j(t) are the phase at time t of the i_{th} and j_{th} hub respectively.

- Phase synchronisation for random assortative connections, the gradient and the base line.

Complete and Phase Synchronisation



(a) Complete synchronization in queue lengths for the 2_{nd} and 3_{rd} CBC hub and (b) Phase synchronization of the queue lengths for certain pair of top most hubs for the gradient mechanism.

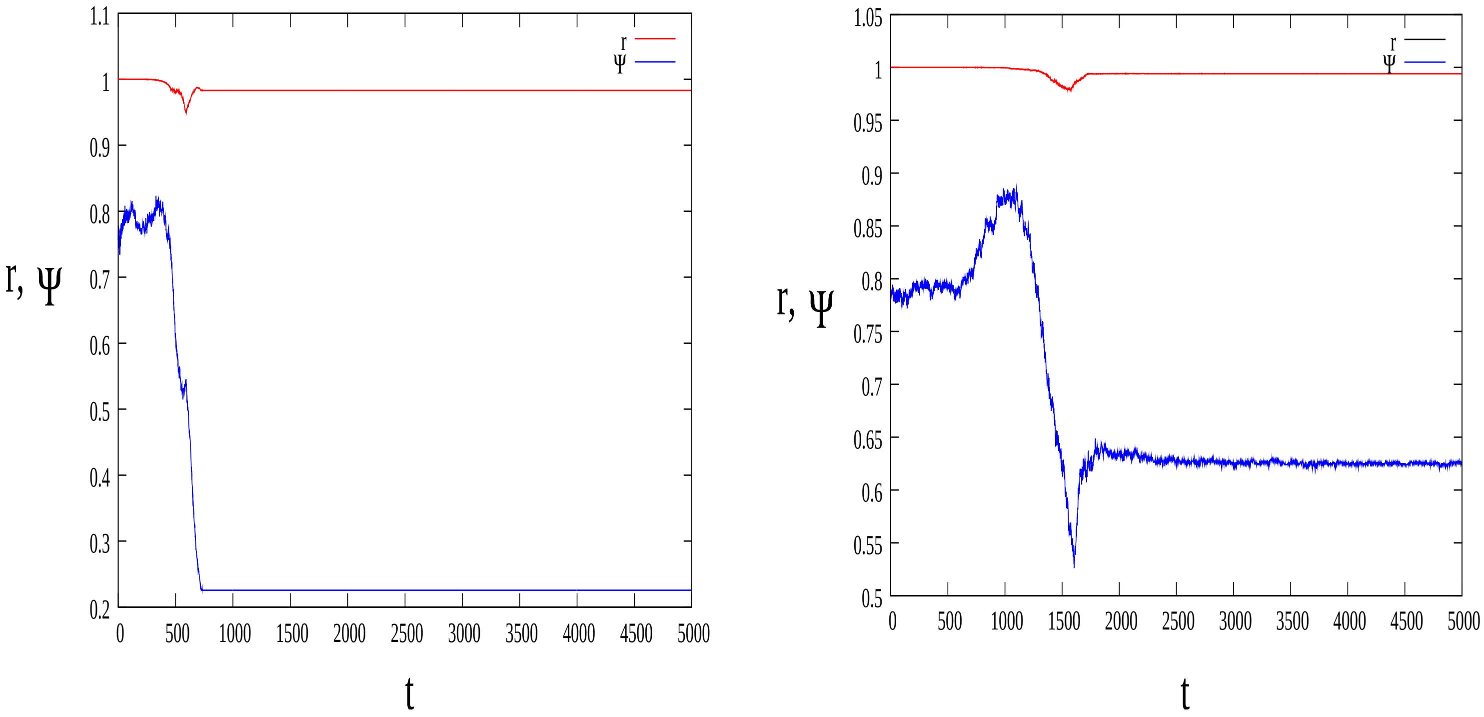
Global synchronisation

- The usual characteriser of global synchronisation is the order parameter

$$r \exp i\psi = \frac{1}{N}\sum_{j=1}^{N} \exp i\Phi_j \qquad (-9)$$

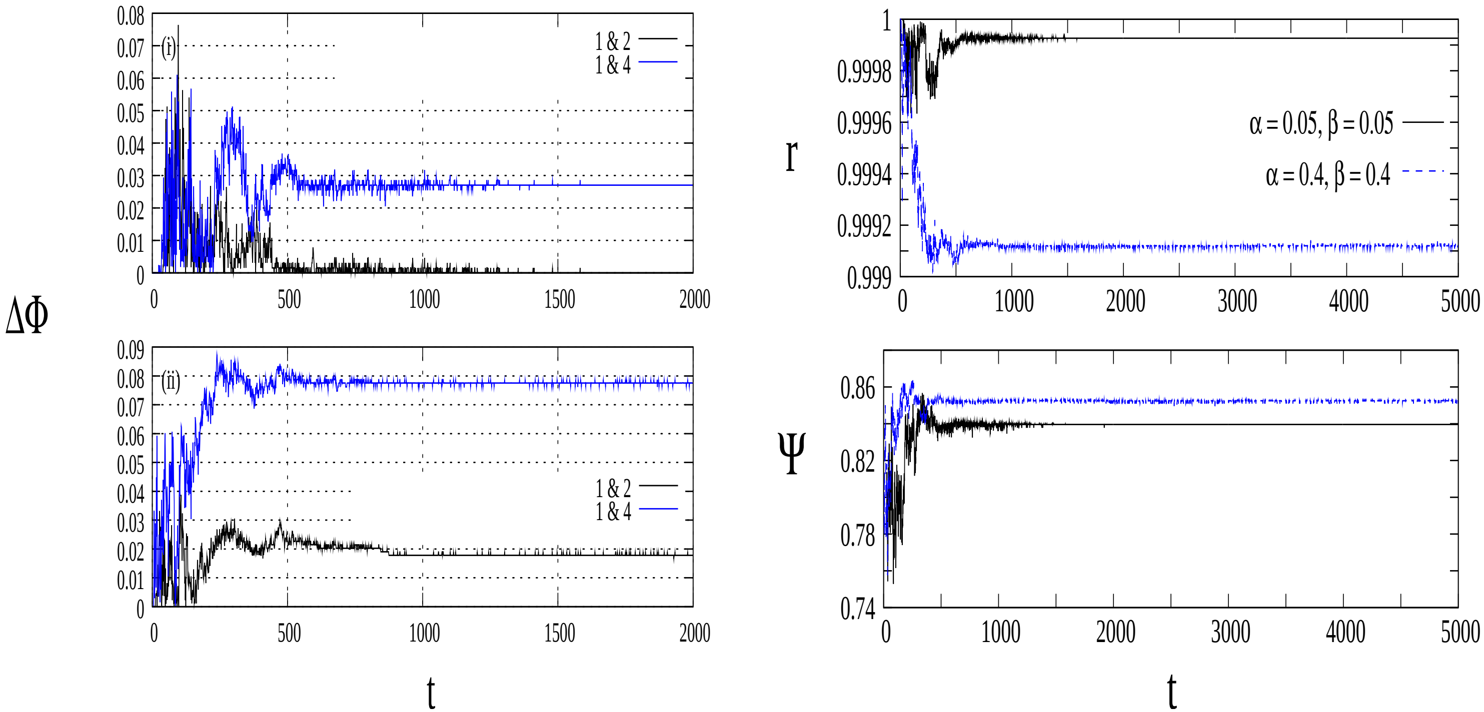
- Here ψ represents the average phase of the system, and the Φ_j-s are the phases defined earlier.
- Here the parameter $0 \leq r \leq 1$ represents the order parameter of the system with the value $r = 1$ being the indicator of total synchronisation.

Global synchronisation



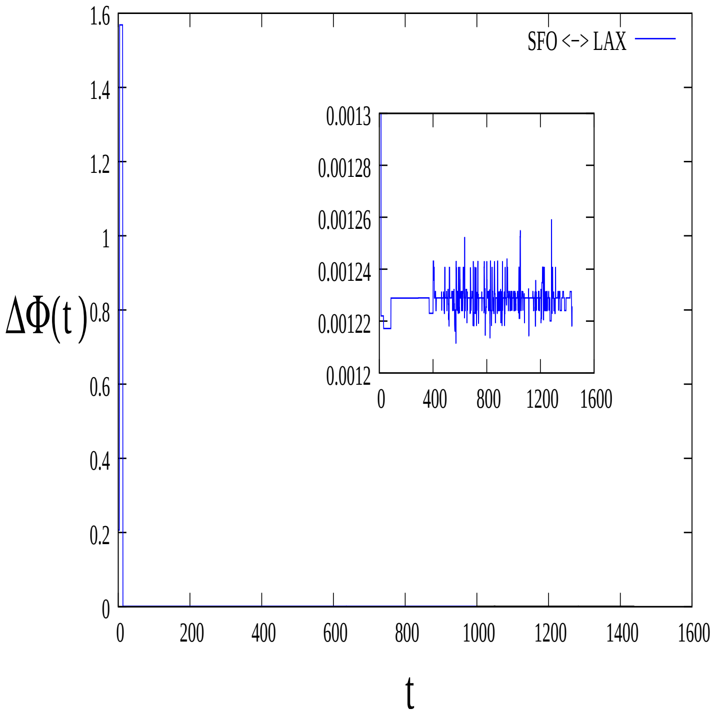
Plot of global synchronization parameters for the baseline mechanism for (a) 2000 messages and (b) 4000 messages flowing simultaneously in the lattice.

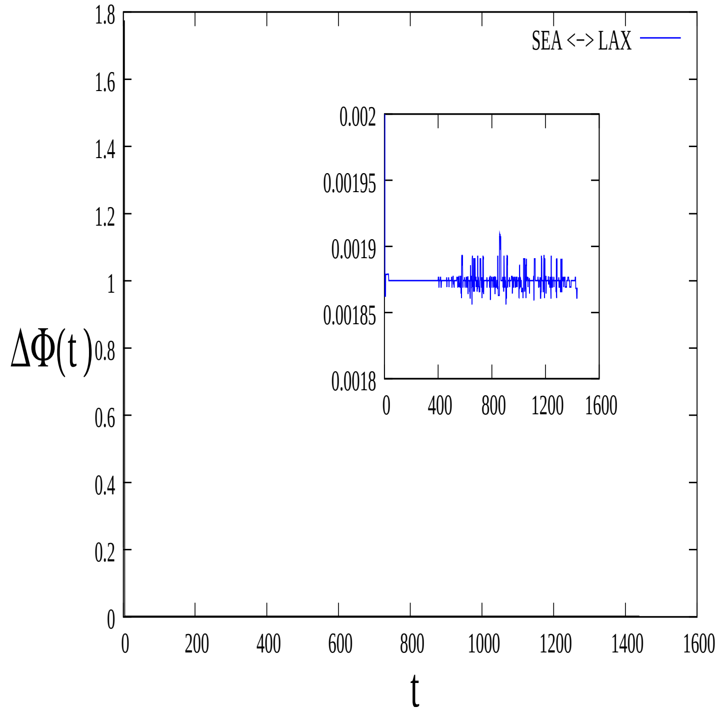
Synchronisation for the Waxman network



Plot of global synchronization parameters for the baseline mechanism for (a) 2000 messages and (b) 4000 messages flowing simultaneously in the lattice.

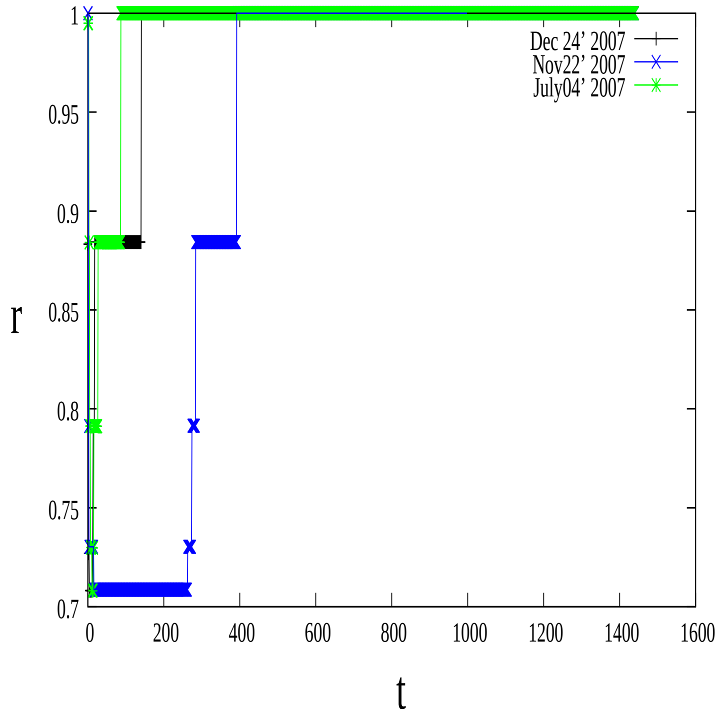
Synchronisation for Airport traffic

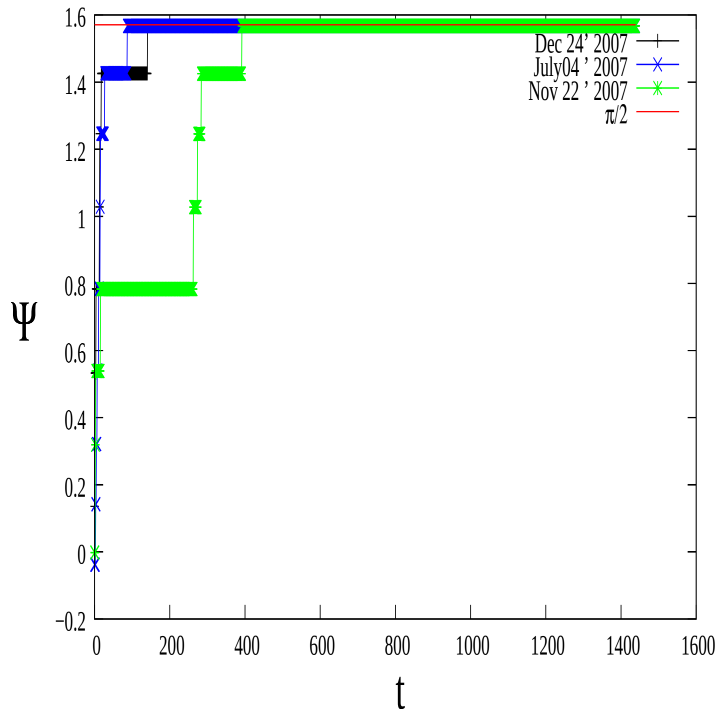




Phase synchronization in queue lengths between (a) SFO and LAX (b) SEA and LAX on Dec. 24 2007.

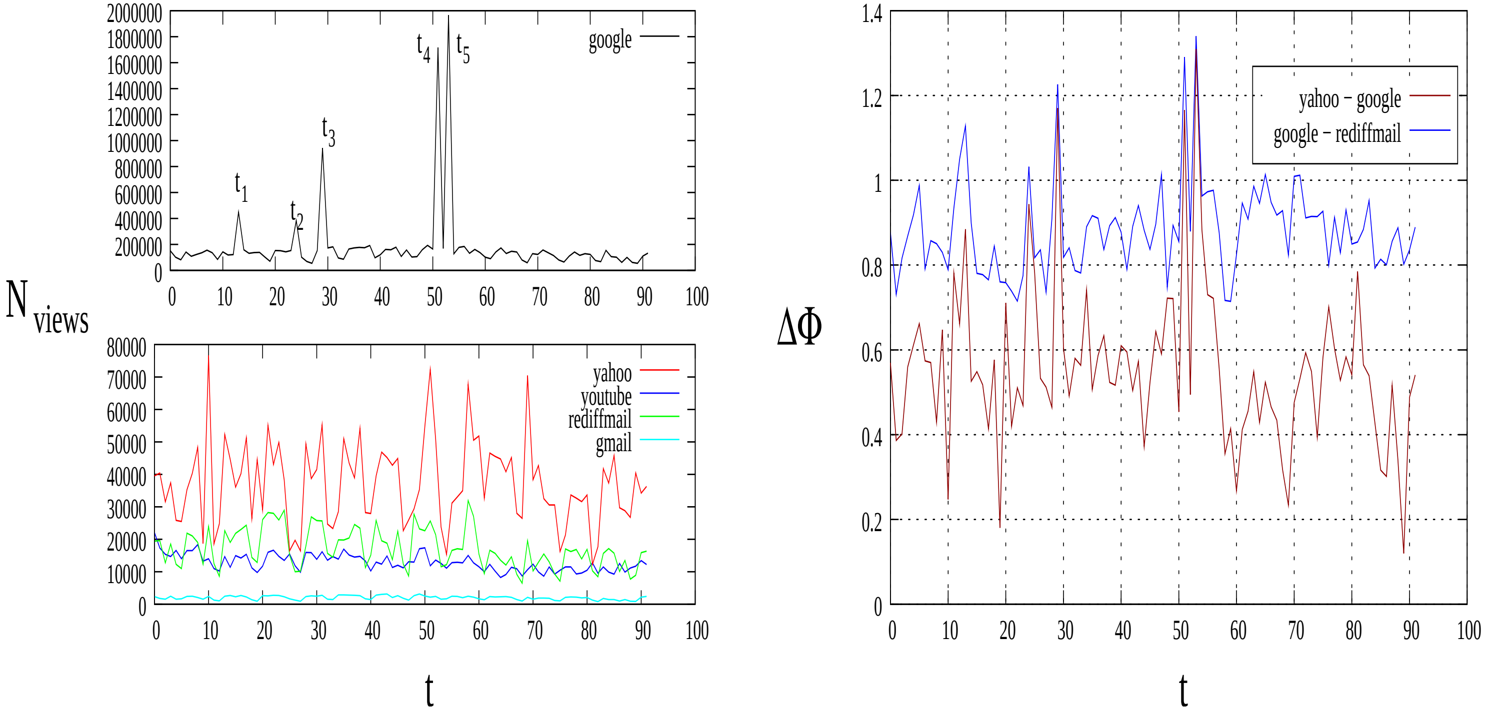
Synchronisation for Airport traffic





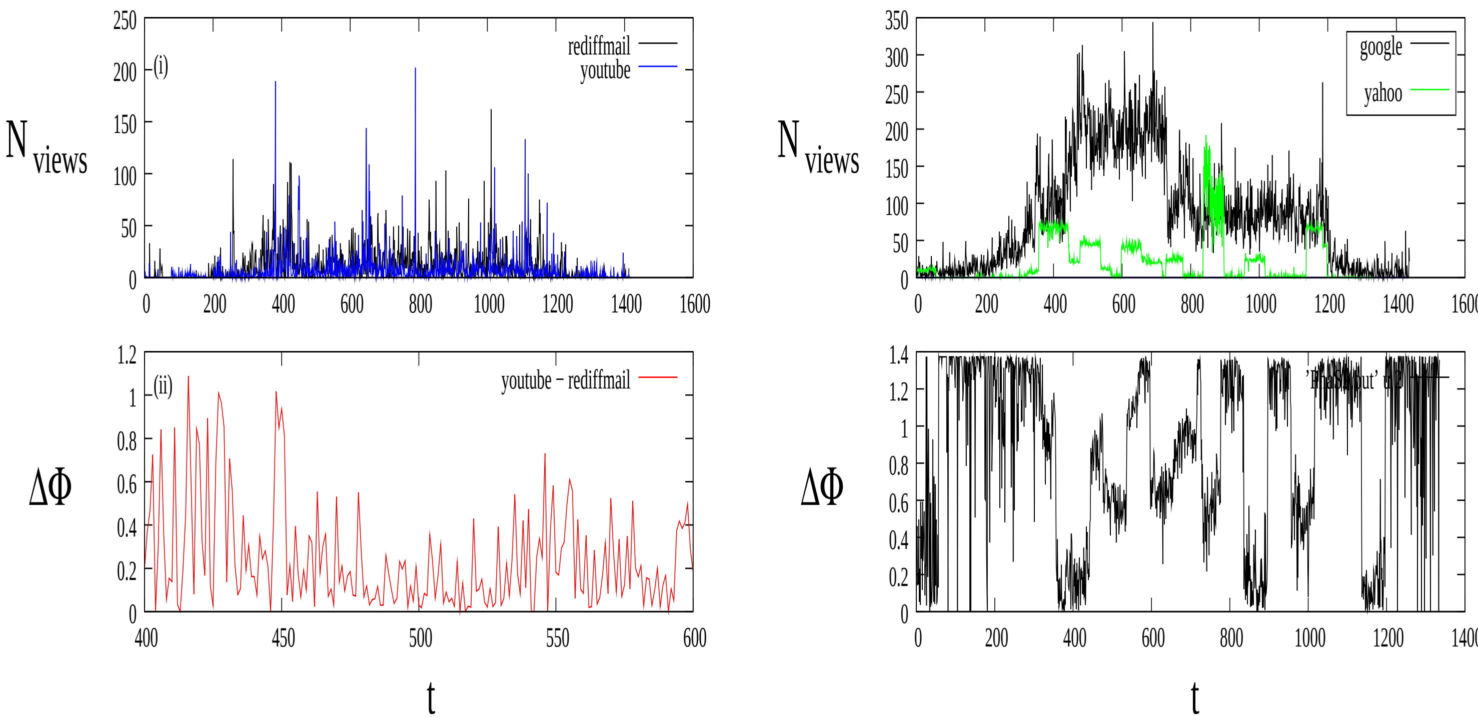
Parameters r ψ t for the USA airport network with 8 airports Los Angeles, San Francisco, Seattle, Miami, Denver, Dallas, Chicago and New York.

Synchronisation for IITM network traffic



No. of views of different web sites for the IIT proxy server between October and December 2nd 2008. Phase synchronisation between pairs of web-sites

Synchronisation for IITM network traffic



Synchronisation effects for the traffic at the IIT proxy server on November 10th

Conclusions

- The statistical characterisers of a several distinct model networks show similar behaviour.
- The average travel times for single message transport show $q-$exponential behaviour as a function of hub density. The power-law tail of this behaviour can be explained in terms of the log-normal distribution of travel times seen at high hub densities.
- The distribution of travel times shows log-normal behaviour for the gradient distribution, and log-normal times powerlaw corrections for the assortative connections.
- In the case of multiple message transfer, the waiting time distribution in the congested phase fits a gaussian. The waiting time distribution in the decongested phase shows log-normal behaviour. This is true across all the networks.

Conclusions

- The queue lengths of the most frequently visited hubs synchronise for similar traffic patterns.This can be complete synchronisation or phase synchronisation. Transitions to total synchronisation can be seen. Synchronisation can therefore be used to detect abnormal traffic. Global synchronisation effects can be used to cluster data.
- Real life networks show similar effects. Our observations can therefore have practical utility.

Thanks to

Andreas Deutschmann (German flying authority), J. Kurths (Univ. of Potsdam), Anil Prabhakar and R.S. Singh (IIT Madras).